



## Breast density classification in mammograms: An investigation of encoding techniques in binary-based local patterns

Bin Rampun, A., Morrow, P.J., Scotney, B., & Wang, H. (2020). Breast density classification in mammograms: An investigation of encoding techniques in binary-based local patterns. *Computers in Biology and Medicine*, 122, [103842]. <https://doi.org/10.1016/j.combiomed.2020.103842>

[Link to publication record in Ulster University Research Portal](#)

**Published in:**  
Computers in Biology and Medicine

**Publication Status:**  
Published (in print/issue): 31/07/2020

**DOI:**  
[10.1016/j.combiomed.2020.103842](https://doi.org/10.1016/j.combiomed.2020.103842)

**Document Version**  
Author Accepted version

**General rights**  
Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**  
The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [pure-support@ulster.ac.uk](mailto:pure-support@ulster.ac.uk).

# Breast Density Classification in Mammograms: An Investigation of Encoding Techniques in Binary-based Local Patterns

Andrik Rampun<sup>a,b,\*</sup>, Philip J. Morrow<sup>b</sup>, Bryan W. Scotney<sup>b</sup>, Hui Wang<sup>c</sup>

<sup>a</sup>*Academic Unit of Radiology, Department of Infection, Immunity and cardiovascular Disease, Sheffield University, S10 2RX, UK*

<sup>b</sup>*School of Computing, Ulster University, Jordanstown, Northern Ireland, BT37 0QB, UK*

<sup>c</sup>*School of Computing, Ulster University, Jordanstown, Newtownabbey, Northern Ireland, BT37 0QB, UK*

---

## Abstract

We investigate various channel encoding techniques when applied to breast density classification in mammograms specifically local binary, ternary and quinary encoding approaches are considered. Subsequently, we propose a new encoding approach based on a seven-encoding technique yielding a new local pattern operator called a local septenary pattern operator. Experimental results suggest that the proposed local pattern operator is robust and outperforms the other encoding techniques when evaluated on the Mammographic Image Analysis Society (MIAS) and InBreast datasets. The local septenary patterns achieved a maximum classification accuracy of 83.3% and 80.5% on the MIAS and InBreast datasets, respectively. The closest comparison achieved by the other local pattern operators is the local quinary operator producing 82.1% (MIAS) and 80.1% (InBreast) maximum accuracies, respectively.

*Keywords:* Breast mammography, Breast Density, Local Binary Patterns, Local Ternary Patterns, Local Quinary Patterns, Local Septenary Patterns

---

---

\*Corresponding author  
Email address: [y.rampun@sheffield.ac.uk](mailto:y.rampun@sheffield.ac.uk) (Andrik Rampun)

## 1. Introduction

Breast cancer is the leading cause of death in women. Although younger women can also get breast cancer, women over 50 years old have a much higher chance of being affected contributing to over 90% of the cases [1]. In 2017, according to the recent breast cancer statistics report by [2], more than 250,000 cases of breast cancer were diagnosed in the United States. Breast Cancer UK revealed that over 11,000 women died in the United Kingdom (UK) in 2016 [3]. Although the causes are not fully understood, several risk factors such as age, family history, overweight/obesity and excessive use of alcohol were found to be the risk contributors of the disease. According to the National Health Services (NHS) in the UK, there are several symptoms of breast cancer such as the appearance of lesions, micro-calcification, breast density and change of size and shape of the breast [4]. Nowadays, mammography is the most common imaging technology used for screening breast cancer to find early symptoms of breast cancer. Other alternative technologies are Ultrasound and Magnetic Resonance Imaging (MRI). In current clinical practice, radiologists have to examine each mammogram of a patient to find any signs of abnormality and doctors can tell whether a tumour is benign or malignant through biopsy tests. Although the overall current clinical methods have significantly improved in the last two decades, there are still some deficiencies such as inconsistency among radiologists, and the fact that it is time-consuming and invasive.

Women with a dense breast are six times more likely to develop cancer mainly because a tumour has a similar appearance to dense tissue, making it difficult to detect in mammograms [5, 6]. Several studies have indicated that there is a significant correlation between breast density and developing breast cancer [5, 6]. In most cases, dense breasts contain more glandular and fibrous tissue. In mammograms, breast regions with dense tissue tend to have more brighter pixels with fatty tissue usually represented by darker pixels. During the screening procedure, breast density estimation is visually assessed by radiologists and classified into four groups based on the amount of dense tissue. The Breast Imaging-Reporting and Data System (BI-RADS) fourth edition is one of the guidelines used in breast density assessment and contains four main categories:

1. BI-RADS I (0-25% dense tissues and mostly fatty)
2. BI-RADS II (26-50% dense tissues with some scattered density)

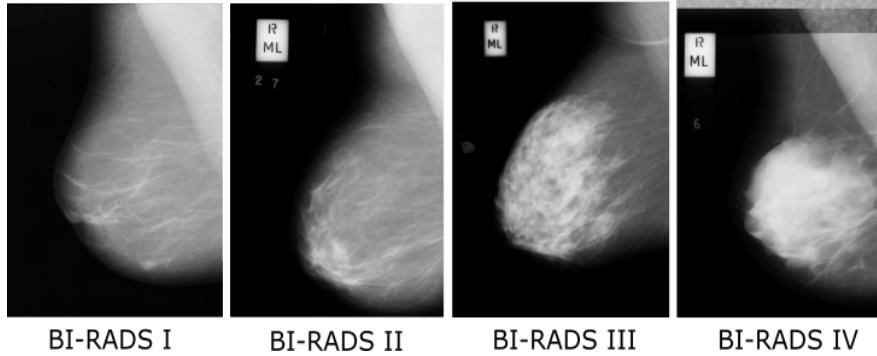


Figure 1: An illustration of breast density according to the BI-RADS guideline fourth edition taken from the the Mammographic Image Analysis Society (MIAS) dataset.

- 37 3. BI-RADS III (51-75% dense tissues and many areas of fibrous and
- 38 glandular tissue that are evenly distributed)
- 39 4. BI-RADS IV (over 75% dense tissues and breasts have a lot of fibrous
- 40 and glandular tissue which makes it hard to see a cancer because it can
- 41 blend in with the normal tissue)

42 Although manual assessment can be done by radiologists, this task is

43 time-consuming and often suffers from variability between radiologists. For

44 example, radiologists with more experience tend to produce more consis-

45 tent results compared to less experienced radiologists. Moreover, manual

46 assessment also could lead to false negatives/positives due to fatigue during

47 diagnostic decision making which could influence the final outcome. Figure

48 1 shows examples of four breasts with their associated BI-RADS classes. In

49 this study we use the fourth edition for simplicity because all datasets are

50 annotated for ground truth based on the BI-RADS fourth edition. However,

51 we are aware that the most recent BI-RADS guidelines is the fifth edition.

## 52 2. Computer-aided diagnosis (CAD) systems

53 Computer-aided diagnosis (CAD) systems can assist clinicians regard-

54 ing efficiency, effectiveness and consistency. CAD systems can assess lesions

55 un-invasively and make predictions as to whether a lesion is benign or malig-

56 nant. Furthermore, CAD can be used as a ‘second reader’ to assist clinicians

57 in diagnostic decision making particularly in cases where doctors are not

58 certain about their decision. However, developing a CAD system that can

replicate radiologists’ knowledge requires a significant amount of time and effort. Machine learning is a sub-component of Artificial Intelligence which enables machines to learn and predict a future occurrence of an event. With the use of machine learning and image processing techniques, it is possible to ‘teach’ computers to learn the appearance/characteristics of breast regions during the training phase and subsequently make a prediction based on the information/knowledge learned.

In breast CAD systems, one of the ways to characterise a breast region in a mammogram is by studying its texture appearance in order to extract meaningful information. In the last decades, many CAD systems for breast imaging have been developed and the majority of them used texture information to investigate the appearance of abnormality such as lesions, dense tissue and micro-calcification clusters. Texture information can be extracted using different mathematical operations such as filter-based, histograms, grey-level distributions, or statistical and probability techniques. Since 2012, the success of deep learning in image classification and segmentation has been overwhelming computer scientists. As a result, hundreds of papers about deep learning for medical image analysis have been published according to the studies by [7] and [8]. Nevertheless, the primary deficiency of deep learning based approaches is the need for a large number of datasets with breast density annotation which can be very challenging in the medical domain. To the best of our knowledge, none of the studies in the literature has attempted using deep learning for four-class breast density classification. The majority of them performed either two-class (low risk (BI-RADS I and II) versus high risk (BI-RADS III and IV)) or three-class (fatty, glandular and dense) classification. On the other hand, many studies have attempted four-class classification using conventional machine learning in conjunction with texture information.

In this study, we are interested in investigating the effects of various encoding techniques in local pattern extraction by dividing the binary patterns into two (local binary patterns (LBP)), three (local ternary patterns (LTP)) and five (local quinary patterns (LQP)) channels. Subsequently, we propose a local septenary patterns (LSP) operator which divides binary patterns into seven channels. Furthermore, we also study the discriminatory level of local patterns extracted from each channel which can help us to determine more informative texture descriptors. The paper has the following contributions:

1. To the best of our knowledge this is the first study attempting to inves-

- 96        tigate the effects of different encoding techniques covering LBP, LTP  
 97        and LQP applied to breast density classification.
- 98        2. We introduce a variant of LBP (two-encoding technique), LTP (three-  
 99        encoding technique) and LQP (five-encoding technique) called local  
 100        septenary patterns (LSP).
  - 101        3. We introduce several ways of improving local patterns' discriminatory  
 102        level in the application to breast density classification via a multichan-  
 103        nel concatenation approach that enables us to combine local binary  
 104        patterns from different channels.
  - 105        4. Whilst threshold values in LTP and LQP need to be set manually by a  
 106        user, we introduce a simple adaptive way to determine threshold values  
 107        in LSP based on the intensity distribution of the neighbourhood.
  - 108        5. Finally we also make a quantitative assessment on each of the channels  
 109        which can reveal more informative texture descriptors.

110        Note that our study does not attempt to develop a new CAD system for  
 111        breast density classification but to investigate the effects on discriminatory  
 112        levels of local patterns using various encoding techniques.

### 113    **3. Breast Anatomy in a Mammogram**

114        Figure 2 shows breast anatomy in a mediolateral-oblique mammogram  
 115        (MLO). In many cases, the upper retroglandular region contains fatty tissue  
 116        and the Corpus Mammarum region contains glandular and dense tissue (if the  
 117        breast is categorised BI-RADS III or IV). Note that usually dense tissue has  
 118        brighter pixels compared to glandular tissue and these tissues have a similar  
 119        appearance to the pectoral muscle region. Since many biological activities in  
 120        the Corpus Mammarum region are due to more Lobules and Ducts, dense tissue  
 121        mostly starts its development in this area.

### 122    **4. Literature Review**

123        Breast density classification in mammograms is one of the most popu-  
 124        lar topics in breast CAD systems together with mass and micro-calcification  
 125        clusters classifications. In a majority of the cases, textures have been a pop-  
 126        ular choice for most authors mainly due to its efficiency and effectiveness.  
 127        Previously, manual and adaptive thresholding techniques have been used to

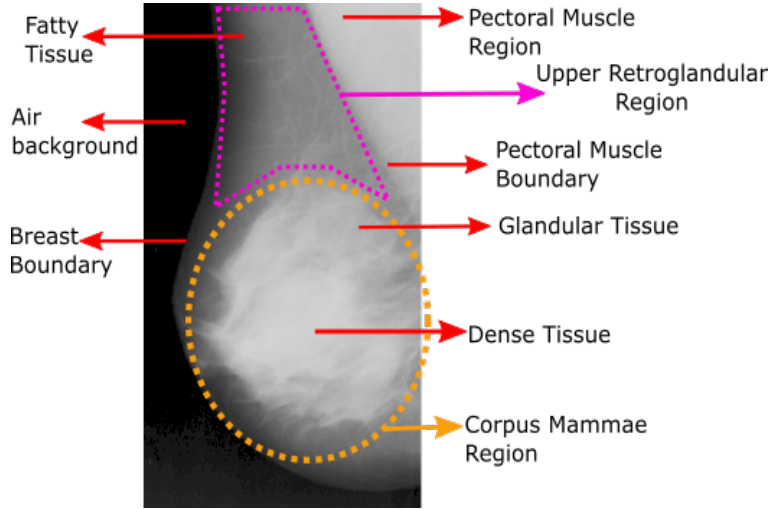


Figure 2: An overview of breast anatomy in a mediolateral-oblique mammogram (MLO). Note that the contrast of the image has been adjusted from the original image for improve visualisation of the breast boundary.

128 estimate the amount of dense tissue. For example, [9] developed an inter-  
 129 active thresholding software called Cumulus to segment regions with dense  
 130 tissue by manually tuning a grey-level threshold value. Due to the difficulty  
 131 in getting annotated mammograms, this topic did not get much attention  
 132 from computer scientists until the MIAS dataset was made available publicly  
 133 in 1994 [10].

134 The first- and second-order (e.g., Grey Level Co-occurrence Matrix (GLCM))  
 135 statistical features are among the most popular texture descriptors. For ex-  
 136 ample the studies of [11], [12], [13] and [14] all used these texture descriptors  
 137 by considering different orientations and resolutions. The methods of [11]  
 138 and [14] first segment the breast into fatty and non-fatty regions followed by  
 139 feature extraction on each region and feature selection to remove uninforma-  
 140 tive descriptors. Several machine learning algorithms were employed by [11]  
 141 and reported an accuracy of 77%, 72% and 86% for  $k$ -Nearest Neighbours,  
 142 C4.5 and Bayesian Network classifiers, respectively. [14] reported an accu-  
 143 racy over 91% using rough fuzzy approaches with the same feature extraction  
 144 techniques used by [11]. [13] selected a set of first- and second-order statis-  
 145 tical features using a combination of several feature selection techniques and  
 146 reported a maximum accuracy of 79.2%. [12] extracted features from the

147 Spatial Grey Level Dependency matrix (a variant of GLCM) as dense tis-  
148 sue descriptors and employed a combined classifier paradigm based on feed-  
149 forward Artificial Neural Network (ANN) models to achieve just over 71%  
150 classification accuracy.

151 Morphology-based descriptors also have been widely investigated to char-  
152 acterise the shape appearance of fatty and non-fatty regions. In many cases,  
153 they are used as a set of descriptors with first- and second-order statistical  
154 features. For example, [11] and [14] extracted a set of morphology features  
155 such as the area, mean, standard deviation and kurtosis using a moment  
156 based histogram technique from the segmented fatty and non-fatty regions.  
157 The fractal-based feature extraction technique based on a set of threshold  
158 values is often used to measure the ratio between the number of low and  
159 high grey-levels within a region of interest (e.g., within a  $5 \times 5$  neighbor-  
160 hood). For example, [15] divided breast regions by recursively splitting it  
161 into quadrants depending on the intensity information that is contained in  
162 each region. Subsequently, each subregion is evaluated until the decision  
163 function determines that the region does not need to be split, thus obtain-  
164 ing regions with similar properties of tissue. They reported an accuracy of  
165 93% based on two-class classification (low versus high risk). [16] used fractal  
166 analysis to characterise parenchymal patterns in digital mammograms for  
167 risk assessment of developing breast cancer, and reported an area under the  
168 curve value of 0.86.

169 Bag-Of-Words features such as Textons and Scale Invariant Feature Trans-  
170 form (SIFT) have also been studied in the last decade. [17] compared the  
171 performance of Textons (extracted from the original images instead of the fil-  
172 tered images) and SIFT features for four-class density classification and used  
173 probabilistic Latent Semantic Analysis (pLSA) to obtain a meaningful set  
174 of features generating a compact tissue representation of each density class.  
175 They reported over 91% and 88% accuracies achieved by Textons and SIFT  
176 approaches, respectively. In a similar study, [18] compared the performance  
177 of various methods for mammographic dense tissue pattern modelling such  
178 as the Local Grey-Level appearance (LGA), Textons, LBP and Basic Image  
179 Features (BIF). In the Textons approach, [18] used various filter sizes and  
180 orientations to extract texture information from filtered images instead of  
181 directly from the original images and various parameter values (e.g., window  
182 size and the number of neighbours) were tested to extract LGA, LBP, and  
183 BIF features. For the four-class BI-RADS classification the authors reported  
184 accuracies 75%, 72%, 59% and 70% for Textons, LGA, LBP and BIF, re-



185 spectively. [19] implemented a Textons approach based on the Maximum  
186 Response 8 (MR8) filter bank. The  $\chi^2$  distribution was used to compare  
187 each of the resulting histograms from the testing set with all the learned  
188 histogram models from the training set and reported 75.5% accuracy.

189 In recent studies, [20, 21] showed that both LTP (three-encoding ap-  
190 proach) and LQP (five-encoding approach) operators can achieve 77% to 85%  
191 accuracy using a multiscale approach with different neighbourhood topolo-  
192 gies and different number of dominant patterns. Instead of extracting texture  
193 features from the entire breast region, [20, 21] showed that extracting texture  
194 information only from the Corpus Mammea reduces the number of overlap-  
195 ping features and hence produces more distinctive features. The authors  
196 showed that this approach was at least 7% better in terms of classification  
197 accuracy compared to classifying breast density by extracting features from  
198 the entire breast region. [22] modelled the distribution of breast density based  
199 on the multi-scale distribution of dark ellipses (representing fatty tissue) and  
200 bright ellipses representing fibroglandular and dense tissues. Subsequently,  
201 morphological features such as distances, sizes and diameters were extracted  
202 to estimate the amount of dense tissue in the breast. Preliminary results  
203 show that their method achieved accuracy between 67% to 72% based on  
204 various classifiers. Later, [23] compared the performance between Elliptical  
205 LBP (ELBP), uniform ELBP and Mean-Elliptical LBP (M-ELBP) for three-  
206 class classification (fatty, glandular and dense) and reported classification  
207 accuracies of 75%, 74% and 80%, respectively.

208 Despite a large number of methods having been developed to address  
209 the breast density classification problem in mammograms, only a few studies  
210 have achieved accuracies above 80% and the majority of the methods pro-  
211 duced between 70% to 79%. Moreover, most studies were based on two-class  
212 classification (low risk (BI-RADS I and II) versus high risk (BI-RADS III and  
213 IV)) or three-class classification (fatty, fatty-glandular and dense-glandular)  
214 instead of four-class classification. This might be due to challenging issues  
215 such as complex and overlapping tissue appearance as well as ambiguous tex-  
216 ture patterns which make it difficult to separate BI-RADS classes based on  
217 texture.

218 In the last a few years, Convolutional Neural Network (CNN) and Deep  
219 Residual Learning (ResNet) are among the most popular deep learning tech-  
220 niques which have been used to address classification and segmentation prob-  
221 lems in clinical image data [8, 7]. Many current studies have claimed that  
222 deep learning based methods produced superior results achieving accuracy

223 similar to human performances [8, 7, 24, 25]. In breast density classifica-  
 224 tion, the majority of the studies used deep learning based methods to address  
 225 either two-class (scattered and heterogeneously density) or three-class clas-  
 226 sification problems (fatty, glandular and dense tissue). [26] developed a  
 227 CNN model to distinguish between scattered density and heterogeneously  
 228 dense tissues using over 22,000 images. [27] used deep learning networks  
 229 to classify dense and non-dense samples for the purpose of dense tissue seg-  
 230 mentation. [28] used unsupervised deep learning to segment dense tissue  
 231 in mammograms and estimate the risk of developing breast cancer based on  
 232 the of segmentation result. Recently, [29] reported that their deep learning  
 233 classification results are correlated well with BI-RADS density assessments  
 234 by radiologists and comparable with a state of the art algorithm, Laboratory  
 235 for Individualized Breast Radiodensity Assessment (LIBRA).

236 Although deep learning based methods have shown promising results,  
 237 breast density classification based on the BI-RADS four-class assessment  
 238 guideline (fourth edition) remains a challenging task. This might be due to  
 239 a lack of annotated data and the complexity of the task itself. According  
 240 to the recent study of [8], the main challenge of employing a deep learning  
 241 network is the requirement of a large dataset and annotations from experts.  
 242 In addition, deep learning suffers from a lack of direct human interpretabil-  
 243 ity because deep learning features rely on filter responses solicited from a  
 244 large amount of training data whereas hand-crafted features such as those  
 245 extracted from local patterns provide transparent information, which are  
 246 more interpretable to clinicians and researchers. Furthermore, all datasets  
 247 available publicly suffer from an imbalance in the number of images for each  
 248 class. For example, in the MIAS [10] and InBreast [30] datasets the number  
 249 of BI-RADS IV cases is less than 10% of the total number of cases. As a  
 250 result, deep learning networks may suffer from an imbalanced classification  
 251 problem.

## 252 5. Methodology

253 This section explains the technical details of our study covering steps  
 254 involved in the pre-processing, feature extraction and the classification phase.  
 255 Figure 3 shows a general overview of the workflow in our experiments.

256 Firstly, we performed breast segmentation to separate the breast bound-  
 257 ary and pectoral muscle using the method proposed by [31]. This ensures  
 258 that only local patterns within the breast region will be extracted. Follow-

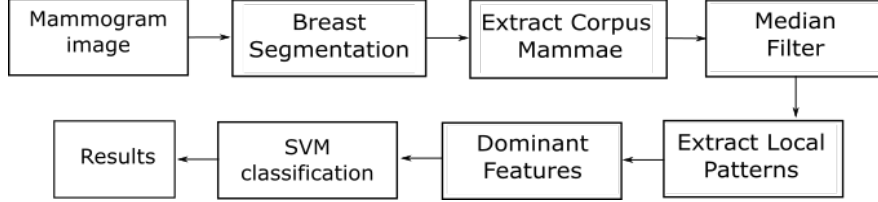


Figure 3: An overview of the work flow involved in our experiments.

259 ing our previous studies [20, 21], we extract the Corpus Mammae region of  
 260 the breast to reduce the risk of overlapping local patterns across BI-RADS  
 261 classes. Subsequently, we used a median filter around  $3 \times 3$  neighbouring pix-  
 262 els for noise reduction. We extract local patterns at different channels using  
 263 the LBP, LTP, LQP and LSP operators within the estimated Corpus Mam-  
 264 mae region. We select a set of dominant patterns to remove uninformative  
 265 features and employ the Support Vector Machine (SVM) as a classification  
 266 approach after a zero-mean normalisation.

### 267 5.1. Pre-processing

268 To segment the breast region, we used our previously developed method  
 269 [31] to estimate the breast and pectoral muscle boundaries. Based on the  
 270 estimated boundaries we created a breast mask and performed a pixel wise  
 271 multiplication with the original image resulting in an image with only the  
 272 breast region. The method [31] employs a region-based Active Contour  
 273 to estimate the breast boundary and restricted contour growing with edge  
 274 information for the pectoral muscle boundary estimation. Figure 4 shows an  
 275 example of an extracted Corpus Mammae region of interest ( $ROI_{cm}$ ) and a  
 276 few examples of estimated breast and pectoral muscle boundaries.

277 The left-most image in Figure 4 shows the estimated  $ROI_{cm}$  area (amber  
 278 square box). To extract  $ROI_{cm}$ , we find the height( $B_h$ ) and the width of the  
 279 breast ( $B_w$ ).  $B_h$  is then relocated to the middle of  $B_w$  to get the intersection  
 280 point. The width and height of the square area of the  $ROI_{cm}$  (amber line  
 281 Figure 4) can be computed as  $B_w \times B_w$  with the centre located at the inter-  
 282 section point between the  $B_h$  and  $B_w$  lines.  $B_h$  is the height of the breast,  
 283 which is the longest perpendicular distance between the  $x$ -axis and the breast  
 284 boundary. Note that the size of  $ROI_{cm}$  varies depending on the width of the  
 285 breast and local patterns are only extracted within  $ROI_{cm}$  instead of from  
 286 the whole breast region as the majority of the studies in the literature have

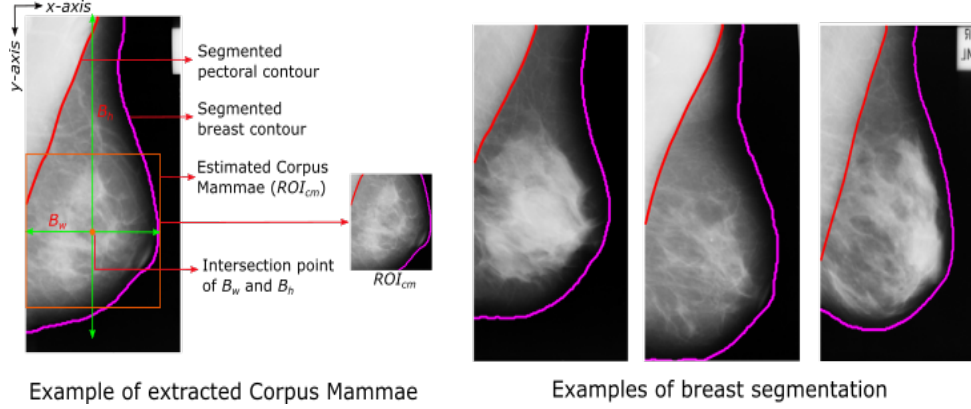


Figure 4: Left: an example of extracted Corpus Mammæ. Right: examples of breast segmentations.

done.

## 5.2. Feature Extraction

This section explains the original LBP operator [32] and its extensions which are the LTP and LQP operators introduced in [33] and [34], respectively. Subsequently, we propose a LSP operator and explain the different channels in each of the operators. Note that there are several other variants of the LBP operator such as ELBP [35], M-ELBP [36], Median Robust Extended Local Binary Pattern (MRELBP) [37], spatio-temporal LBP (STLBP) [38] and completed LBP (CLBP) [39]. However, they are outside the scope of our study because these operators are based on the LBP's two-channel encoding technique (nevertheless we consider this as part of our future work) whereas the scope of our study is to investigate the effects of local patterns on accuracy when various different channel encoding techniques are employed.

All these operators use three different mapping tables namely uniform pattern ('u2'), rotation invariant pattern ('ri') and both uniform and rotation invariant patterns ('riu2') to extract uniform, rotation invariant and rotation invariant uniform patterns, respectively. In our study, we investigate the 'riu2' patterns (a combination of 'u2' and 'ri' patterns) as they provide more discriminant features based on our previous studies [20, 21] and the study of

307 [32]. The value of the LBP code of the pixel  $(i, j)$  is given by:

$$LBP_{(P,R)}(i, j) = \sum_{p=0}^{(P-1)} s(g_p - g_c) 2^p \quad (1)$$

308 where  $R$  and  $P$  are the radius of the circle that forms the neighbourhood of  
 309 the operator and the number of pixels in the neighbourhood, respectively.  
 310 The grey level value of the centre pixel is denoted  $g_c$ , and  $g_p$  is the grey level  
 311 value of the  $p^{th}$  neighbour. The LBP operator thresholds the neighbouring  
 312 pixels using a two-value encoding system as shown in Equation 2.

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

313 Later, [33] introduced a three-value encoding technique (LTP operator)  
 314 which thresholds the neighbouring pixels based on a constant threshold set by  
 315 the user ( $\tau_1$ ) as shown in Equation 3. Once the LTP code image is generated,  
 316 it can be separated into two binary patterns from its positive and negative  
 317 channels. Therefore, it encodes an image into three channels but producing  
 318 two binary patterns.

$$s(x) = \begin{cases} 1, & x > g_c + \tau_1 \\ 0, & g_c - \tau_1 < x < g_c + \tau_1 \\ -1, & x < g_c - \tau_1 \end{cases} \quad (3)$$

319 In [34] the LQP operator was introduced, which thresholds the neigh-  
 320 bouring pixels using a five-value encoding technique (see Equation 4) based  
 321 on two constant thresholds  $\tau_1$  and  $\tau_2$ . Subsequently, the LQP code image is  
 322 split into four binary patterns by considering its positives, zero and negative  
 323 components. Therefore, the LQP operator encodes an image into five chan-  
 324 nels but results four binary patterns (from two positive channels ( $s(1)$  and  
 325  $s(2)$ ), zero channel ( $s(0)$ ) and combined negative channels ( $s(-1) \cap s(-2)$ )).

$$s(x) = \begin{cases} 2, & x \geq g_c + \tau_2 \\ 1, & g_c + \tau_1 \leq x < g_c + \tau_2 \\ 0, & g_c - \tau_1 \leq x < g_c + \tau_1 \\ -1, & g_c - \tau_2 \leq x < g_c - \tau_1 \\ -2, & \text{otherwise} \end{cases} \quad (4)$$

Figure 5 shows the resulting code images and local patterns from each channel using the LBP, LTP and LQP operators. A histogram from the LBP code image ( $H_{lbp}$ ) can be extracted to characterise local patterns in  $ROI_{cm}$ . However, for LTP and LQP code images, local patterns can be extracted from each channel. For example, local patterns that represent  $ROI_{cm}$  when using the LTP operator can be extracted by computing the histogram image of ‘LTP Channel 1’ and ‘LTP Channel 2’, donated as  $H_{ltp1}$  and  $H_{ltp2}$ , respectively and concatenating these histograms as the final feature descriptors. Similarly, to extract local patterns in  $ROI_{cm}$  using LQP operator, we can compute and concatenate the histograms from all LQP channels ( $H_{lqp1}$ ,  $H_{lqp2}$ ,  $H_{lqp3}$  and  $H_{lqp4}$ ). Note that all channel images are binary images. As a result of histogram concatenating more local patterns can be extracted using the LTP and LQP operators.

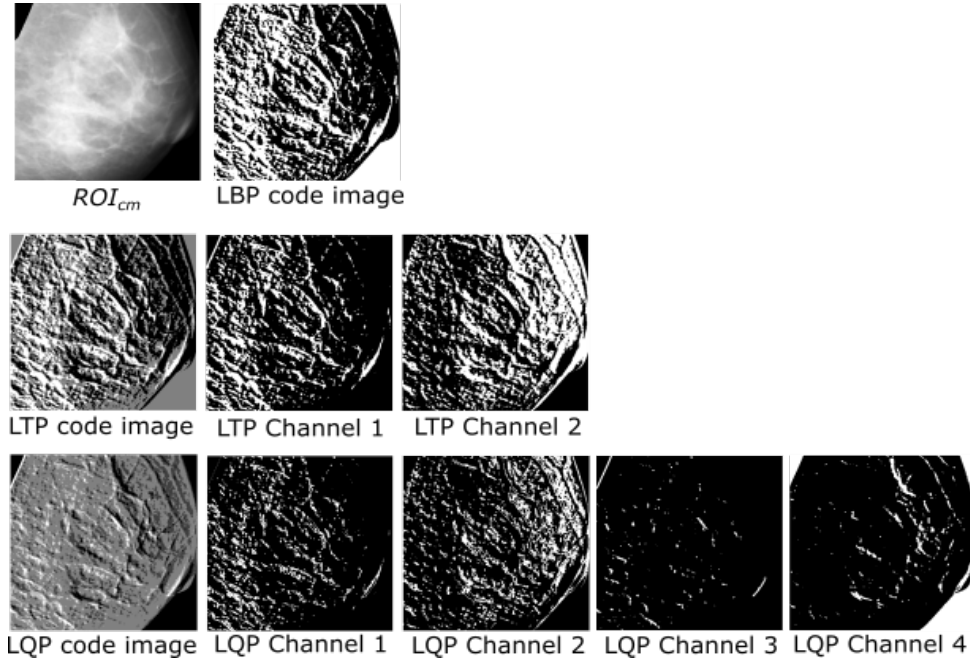


Figure 5: Examples of LBP, LTP and LQP code images and local pattern images from different channels.

We took a similar approach by thresholding the neighbouring pixels using a seven-value encoding technique (hence the name is septenary) based on three threshold values ( $\tau_1$ ,  $\tau_2$  and  $\tau_3$ ). The value of the LSP code of the pixel

342  $(i, j)$  is given by:

$$LSP_{(P,R)}^{pattern}(i, j) = \sum_{p=0}^{(P-1)} s_{pattern}(g_p) 2^p \quad (5)$$

343 where  $pattern \in \{1, 2, 3, 4, 5, 6\}$  represents six binary patterns by considering  
 344 its upper-positive, middle-positive, lower-positive, upper-negative, middle-  
 345 negative and lower-negative components denoted as  $s(3)$ ,  $s(2)$ ,  $s(1)$ ,  $s(-1)$ ,  
 346  $s(-2)$  and  $s(-3)$ , respectively, as shown in Equation 6

$$s(x) = \begin{cases} 3, & x \geq g_c + \tau_3 \\ 2, & g_c + \tau_2 \leq x < g_c + \tau_3 \\ 1, & g_c + \tau_1 \leq x < g_c + \tau_2 \\ 0, & g_c - \tau_1 \leq x < g_c + \tau_1 \\ -1, & g_c - \tau_2 \leq x < g_c - \tau_1 \\ -2, & g_c - \tau_3 \leq x < g_c - \tau_2 \\ -3, & \text{otherwise} \end{cases} \quad (6)$$

347 Note that LSP encodes an image into seven channels but results in six binary  
 348 patterns. The LSP code image is split into six binary patterns using the  
 349 following conditions

$$s_1(x) = \begin{cases} 1, & \text{if } s(3) = 3 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$s_2(x) = \begin{cases} 1, & \text{if } s(x) = 2 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

$$s_3(x) = \begin{cases} 1, & \text{if } s(x) = 1 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

$$s_4(x) = \begin{cases} 1, & \text{if } s(x) = -1 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

$$s_5(x) = \begin{cases} 1, & \text{if } s(x) = -2 \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

$$s_6(x) = \begin{cases} 1, & \text{if } s(x) = -3 \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

Figure 6 shows an example of a LSP code image extracted from  $ROI_{cm}$  in Figure 5 and its local patterns from different channels. Using the encoding technique in Equation 6, we can generate the LSP code image of  $ROI_{cm}$ . The LSP code image is split into six binary images which represent local patterns in six channels. Unlike LTP and LQP where the user has to manually determine threshold values, we introduce an automatic approach by computing the number of neighbours with grey level value  $\leq 25^{th}$  percentile of the entire neighbourhood, number of neighbours with grey level value between  $25^{th}$  and  $75^{th}$  percentile of the entire neighbourhood and number of neighbours with grey level value  $\geq 75^{th}$  percentile. Subsequently, we sort the values in ascending order for  $\tau_1$ ,  $\tau_2$  and  $\tau_3$  values.

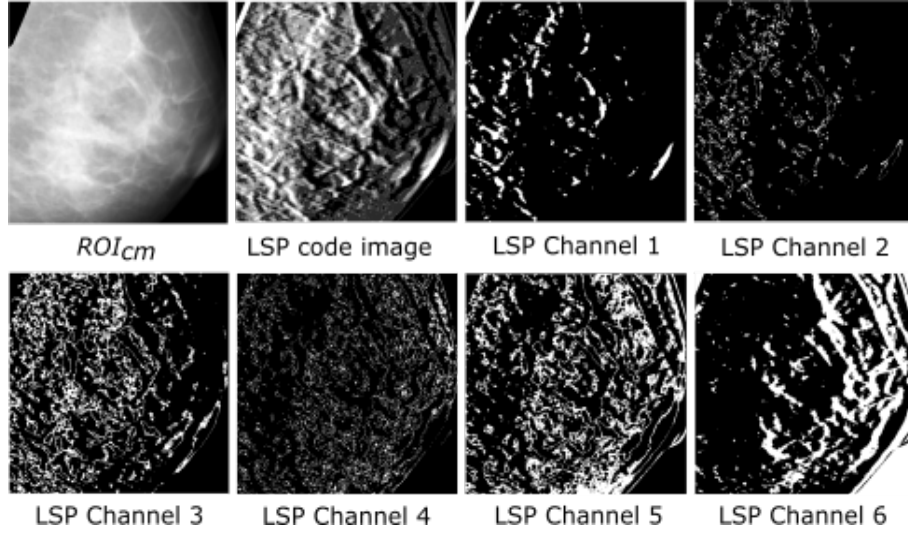


Figure 6: Examples of LSP code image and its local pattern images from different channels.

The LBP, LTP, LQP and LSP are similar in terms of architecture as each is defined using a circle centred on each pixel and the number of neighbours. The main difference is the LSP thresholds the neighbouring pixels into seven (3, 2, 1, 0, -1, -2 and -3) values. In contrast, the LBP, LTP and LQP encode an image into two (1 and 0), three (1, 0 and -1), and five (2, 1, 0, -1 and -2) values. Figure 7 shows an example of the feature extraction process. The final



367 histogram is a concatenation of six histograms computed from each binary  
 368 pattern (binary image, e.g., LSP code 1) generated based on the conditions  
 369 in equations (7) to (12). The process of capturing local patterns of  $ROI_{cm}$  is  
 370 also the same for the other operators such as the LBP, LTP and LQP, except  
 371 less number of channels (resulting in a shorter concatenated histogram).

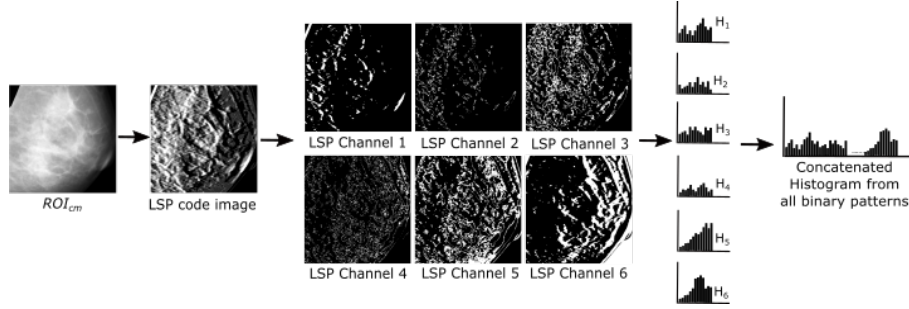


Figure 7: Summary of feature extraction phase using the LSP operator. Note that the binary pattern from each channel is computed from the LSP code image and a histogram image is computed for each channel. Finally, all histograms are concatenated and treated as a feature vector to represent the local pattern of  $ROI_{cm}$ .

### 372 5.3. Dominant Patterns

373 Following our previous studies in [21, 20], we selected dominant patterns  
 374 to reduce feature dimension, hence simplifying the prediction model. Domi-  
 375 nant patterns are patterns that occur frequently in the training set [40]. Let  
 376  $I_1, I_2 \dots I_j$  be images in the training set. We compute the histogram feature  
 377 ( $H_{I_j}^{LSP}$ ) for each training image and perform a bin-wise summation for all  
 378 the histograms to find the pattern's distribution from the training set. Sub-  
 379 sequently, the resulting histogram ( $H^{LSP}$ ) is sorted in descending order, and  
 380 the patterns corresponding to the first  $D$  bins are selected using the following  
 381 equation:

$$D = \arg \min_N \frac{\sum_{i=1}^{N-1} H^{LQP}(i)}{\sum_{i=1}^{2^P} H^{LQP}(i)} > 0.01 \times n \quad (13)$$

382 where  $N$  and  $n$  are the total number of patterns and the threshold chosen  
 383 by the user, respectively. For example choosing  $n = 98$  means removing  
 384 patterns that occur with relative frequency of less than 2% in  $H^{LSP}$ . The  
 385 smaller the value of  $n$ , the smaller the number of patterns selected.

#### 386 5.4. Classification

387 Once the feature extraction is completed, we employed several machine  
 388 learning algorithms in WEKA [41] to build our classification models. For  
 389 machine learning with only one parameter (e.g.,  $k$ -NN), the CVPParameterS-  
 390 election technique was employed. In contrast, the GridSearch technique was  
 391 used to explore two parameters for classifiers with two parameters (e.g. RF  
 392 and SVM). The CVPParameterSelection finds the best setting based on op-  
 393 tions set by the user by optimising the classification accuracy. The Grid-  
 394 Search tests all possible combination of two parameters and selects the best  
 395 combinations based on the highest accuracy. To find the best parameters for  
 396 each classifier, 25 patients (9, 7, 5, 4 patients for BI-RADS I, II, III and IV,  
 397 respectively) were selected from the MIAS dataset and 3-fold cross-validation  
 398 was used to evaluate the performance for each (or pair) of the tested param-  
 399 eter(s) during the optimisation process. The classifier was trained, and in  
 400 the testing phase, each unseen  $ROI_{cm}$  from the testing set is classified as  
 401 BI-RADS I, II, III or IV. The classifiers used in this study are presented in  
 402 Table 1.

Table 1: List of classifiers used in this study. The # sign represents number

Classifiers	Default parameters in WEKA
Support Vector Machine (SVM)	Complexity ( $C$ )=1.0, exponent=1
Random Forest (RF)	# of forests=100, depth = 0 (unlimited)
Multilayer Perceptron (MLP)	Learning rate=0.3, momentum=0.2
$k$ -Nearest Neighbours ( $k$ -NN)	$k$ =1, Euclidean distance

## 403 6. Experimental Results

404 To test the performance of the method we used the following datasets:  
 405 (a) the Mammographic Image Analysis Society (MIAS) dataset [10] which  
 406 consists of 322 mediolateral-oblique (MLO) mammograms of 161 women.  
 407 Films were taken from the UK National Breast Screening Programme and  
 408 have been digitised to  $50\mu\text{m} \times 50\mu\text{m}$  and quantised to 8 bits. The distribution  
 409 for BI-RADS classes is 60 (BI-RADS I), 105 (BI-RADS II), 129 (BI-RADS  
 410 III) and 31 (BI-RADS IV), and (b) the InBreast dataset [30] which consists  
 411 of 206 MLO mammograms from 103 patients. Each image is direct digital  
 412 acquisition on a-Se imaging plates. The pixel size of all images is  $70\mu\text{m} \times$

70  $\mu\text{m}$ , with 14-bit contrast resolution. The density distribution for the BI-RADS classes is as follow: 69 (BI-RADS I), 74 (BI-RADS II), 49 (BI-RADS III), and 14 (BI-RADS IV). Each image contains BI-RADS information (e.g., BI-RADS class I, II, III or IV) provided by an expert radiologist based on the fourth edition of the BI-RADS system. We ran stratified 5-fold cross validation for 10 times.

Accuracy ( $Acc$ ) is used to measure the performance of the method, which represents the total number of correctly classified images as a proportion of the total number of images.

### 6.1. Optimised Parameters

Table 2 shows a list of parameter values tested and optimised values for the four classifiers employed in this study. Note that, the parameter optimisation was performed based on 25 patients (each BI-RADS class has six or seven patients) taken from the MIAS dataset. Subsequently, we use these parameter values in the testing phase for both MIAS and InBreast datasets. However, for the MIAS dataset we only use 136 patients (272 mammograms) and we excluded 25 patients (50 mammograms) which were used for parameter optimisation. From now on, all parameter settings for the classifiers employed in this study are based on the best parameter values summarised in Table 2.

Table 2: List of parameter options tested in this study.

Classifiers	Parameter tested	Best parameters
SVM	Kernel = 'Polynomial', $C = 1$ to 10 $e = 1$ to 5	$C = 5$ , $e = 1$
RF	Number of forests ( $rF$ )=1 to 165 depth ( $D$ )=0 to 10	$rF = 70$ , $D = 0$
MLP	Learning rate ( $LR$ )=0.1 to 2.0 (e.g. 0.1, 0.2 ... 2.0) momentum( $M$ )=0.1 to 2.0 (e.g. 0.1, 0.2 ... 2.0)	$LR = 0.1$ , $M = 0.5$
$k$ -NN	$k$ =1-49 (e.g. 1, 3, 5, ... 49) Euclidean distance	$k = 5$

### 6.2. Quantitative Results

This section presents classification results for LBP, LTP, LQP and LSP operators based on different classifiers. Since the LTP and LQP require

threshold values from a user, the parameters  $\tau_1$  and  $\tau_2$  were selected based on previous studies [21, 20]. Therefore the  $\tau_1$  value for the LTP operator is set to 5 and  $\tau_1$  and  $\tau_2$  values for LQP are set to 5 and 12, respectively. Note that the average *Acc* represents the mean accuracy across different numbers of dominant patterns ( $n$ ) where the maximum *Acc* is the highest accuracy achieved for  $n$  number of dominant patterns. Choosing  $n = 93$  means removing patterns that occur with a frequency of less than 7% in the histogram feature. Numerical values represent performance evaluation based on 272 and 206 images for the MIAS and InBreast datasets, respectively. Note that we have used 50 images (of the original 322 images) from the MIAS dataset for parameter optimisation.

Figure 8 and 9 shows classification results using LBP operator for the MIAS and InBreast dataset, respectively. It can be observed that for the MIAS dataset a maximum accuracy of 73.8% is achieved by  $LBP(7, 16)$  (where  $R = 7$ ,  $P = 16$ , e.g.  $LBP(R, P)$ ) using the SVM classifier at  $n = 91.5$  (removing local patterns with frequency of less than 8.5% in the histogram features). The best average accuracy of 69.7% is achieved when employing  $LBP(7, 16)$  with  $n = 92.1$ . For the InBreast dataset a maximum accuracy of 73.8% is achieved by  $LBP(5, 10)$  at  $n = 96.5$  whereas best average accuracy (70.2%) is achieved by  $LBP(9, 20)$  using the MLP classifier. Overall, the majority of the classification accuracies fall in a range between 65% to 70%.

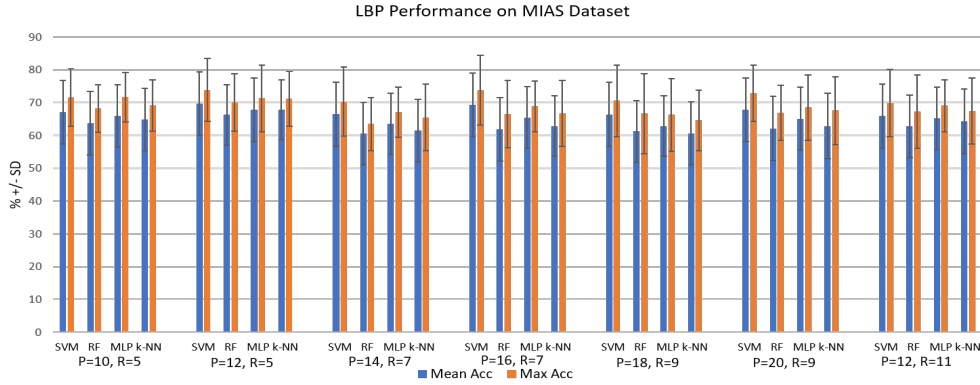


Figure 8: LBP performance on MIAS dataset

Figure 10 and 11 shows quantitative results using the LTP operator for the MIAS and InBreast dataset, respectively. The majority of the average accuracies are in a range between 70% to 75% which are higher than the

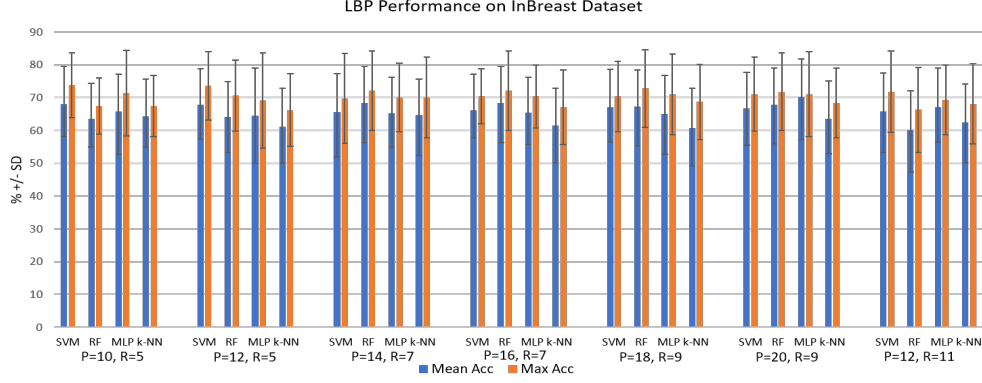


Figure 9: LBP performance on InBreast dataset

ones produced by the LBP operator; for maximum accuracy, many cases are over 75%. Furthermore, it can be observed that the LTP operator outperformed the LBP operator regardless of the  $R$  and  $P$  values. The highest accuracies are 81% (average 77.5%) and 78.7% (average 74.3%) on the MIAS and InBreast datasets, respectively. They are at least 4% to 7% higher than the maximum accuracy produced when extracting local patterns using the LBP operator. The SVM classifier produced the highest classification accuracy by removing 6.5% of the local patterns when evaluated on the MIAS dataset ( $LTP(7, 14)$  and  $LTP(7, 16)$ ) where the MLP classifier ( $LTP(9, 18)$ ) outperformed the other classifiers at  $n = 93.1\%$  when tested on the InBreast dataset.

Figure 12 and 13 shows classification results using the LQP operator for the MIAS and InBreast dataset, respectively. Overall, it can be observed that LQP produced better classification results in comparison to LBP and LTP operators. The majority of the maximum accuracies are in a range between 74% to 79% which indicates that local patterns extracted by the LQP operator are more discriminant in comparison to the previous two operators.  $LQP(9, 18)$  and  $LQP(9, 20)$  produced a maximum accuracy of 82.1% with the best average accuracy of 78.6% at  $n = 90$  on the MIAS dataset. However, when evaluated on the InBreast dataset the maximum accuracy of 80.1% is achieved by  $LQP(9, 18)$  at  $n = 98$  and the best average accuracy is achieved by  $LQP(9, 20)$ . Moreover, it can also be observed that the SVM classifier produced an average maximum accuracy over 80% regardless of the values of  $P$  and  $R$  when evaluated on the MIAS dataset. However, the MLP classifier

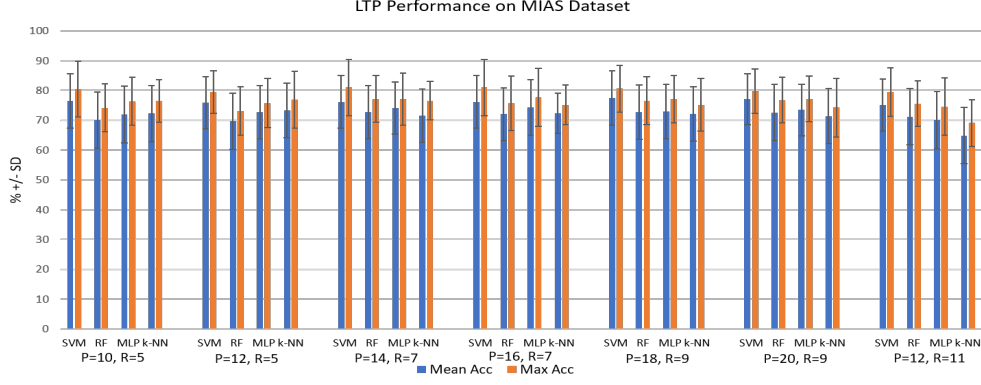


Figure 10: LTP performance on MIAS dataset

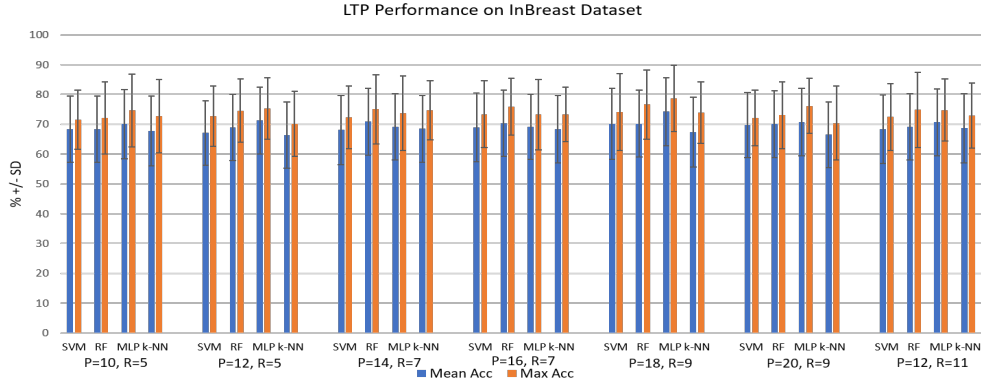


Figure 11: LTP performance on InBreast dataset

484 tends to produce better results when tested on the InBreast dataset with the  
 485 best average  $Acc = 75.5\%$  which is 1% better than the SVM classifier.

486 Table 14 and 15 shows quantitative results using our proposed LSP op-  
 487 erator when evaluated on the MIAS and InBreast dataset, respectively. The  
 488 majority of the classification accuracies fall in the range 75% to 80% which  
 489 is slightly better compared to the results produced by the LQP operator.  
 490 Experimental results show that the LSP operator produced a maximum accu-  
 491 racy of 83.3% (which outperforms the LBP (73.8%), LTP (81%) and LQP  
 492 (82.1%) operators) on the MIAS dataset. The best average accuracy of 81.6%  
 493 is achieved by  $LSP(9, 18)$  at  $n = 99.1$  which is at least 2.8% higher than the

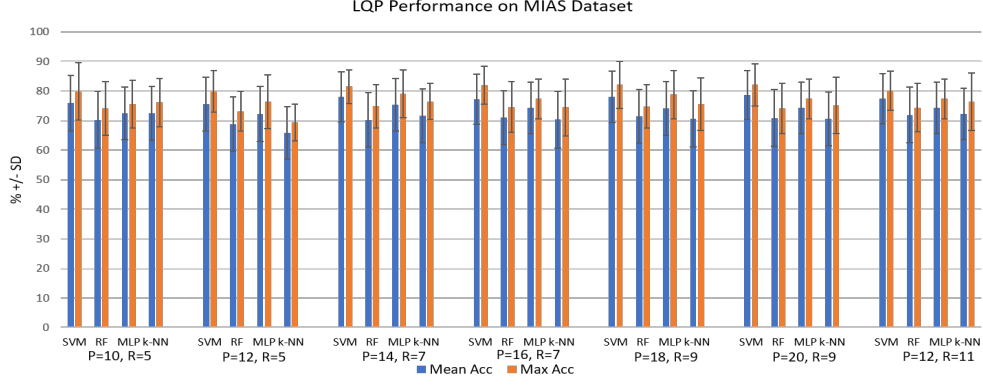


Figure 12: LQP performance on MIAS dataset

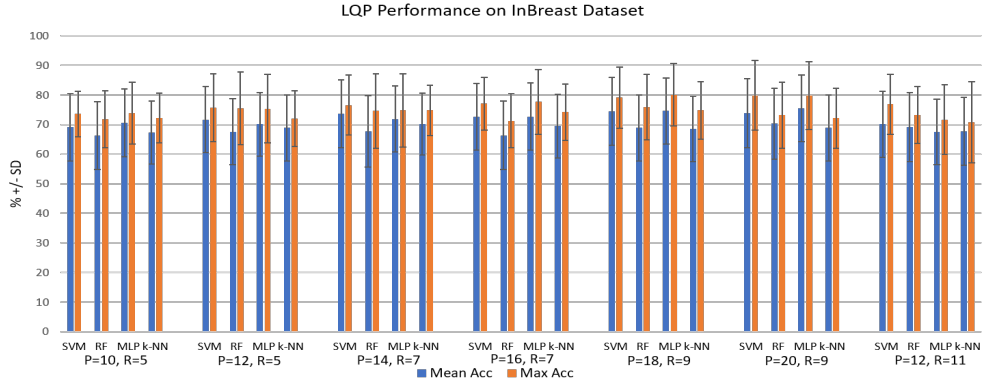


Figure 13: LQP performance on InBreast dataset

best average  $Acc$  produced by LQP and LTP. The LSP operator achieved maximum 80.5% classification accuracy at  $n = 95.5$  using the SVM classifier whereas the highest accuracy achieved by RF,  $k$ -NN and MLP are 77.9 ( $LSP(11, 22)$ ), 79.7% ( $LSP(9, 20)$ ) and 79.8 ( $LSP(7, 16)$ ), respectively. Regarding the best average accuracy across different numbers of dominant patterns  $LSP(9, 20)$  produced 77.1% using the SVM classifier. RF,  $k$ -NN and MLP produced 70.9%, 73.3% and 73.7%, respectively when  $LSP(9, 20)$  is employed. Overall, the LSP operator produced more discriminant local patterns in separating the BI-RADS classes in both MIAS and InBreast datasets. We will discuss this further in the subsequent subsections.

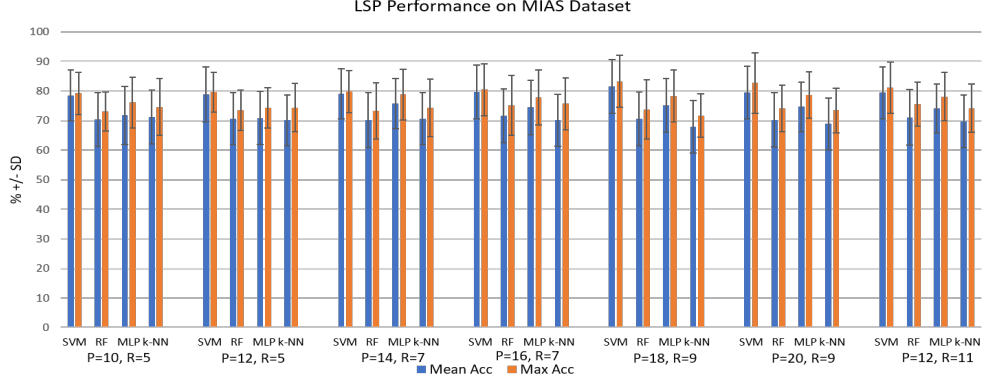


Figure 14: LSP performance on MIAS dataset

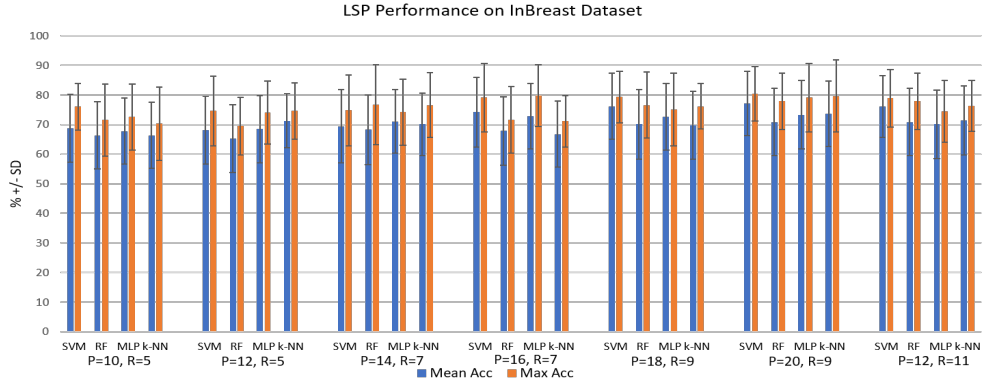


Figure 15: LSP performance on InBreast dataset

### 6.3. Maximum and Average Performance Comparison Between Operators

This section summarises the maximum and average results produced by the operators across different classifiers based on the MIAS and InBreast datasets. For the MIAS dataset, Figures 16 and 17 show performance comparisons between LBP, LTP, LQP and LSP operators using different classification approaches. In terms of maximum accuracy, it can be observed that LSP ( $Acc = 83.3\%$ ) outperformed the other texture operators (LBP, LTP and LQP) when using SVM, RF and MLP as classification approaches. However, when the  $k$ -NN classifier is employed, the maximum classification accuracy is  $76.9\%$  using LTP which is  $0.7\%$  and  $1.2\%$  better than the LQP



and LSP operators, respectively. Regarding the average accuracies across different numbers of dominant patterns, both the SVM and MLP classifiers produced higher results when using LQP and LSP features. For example, the LSP operator produced 81.6% and 75.8% accuracies for SVM and MLP classifiers, respectively compared to 78.6% (SVM) and 75.4% (MLP) when classification was performed based on local patterns extracted using the LQP operator. In contrast, classification accuracy is at least 0.9% better for local patterns using LTP when RF and  $k$ -NN classifiers are used compared to LQP and LSP. In both Figures 16 and 17, the LBP operator produced the lowest accuracy results regardless to classification approach which clearly shows that a channel encoding technique is necessary to capture more discriminant features and hence improve the classification results.

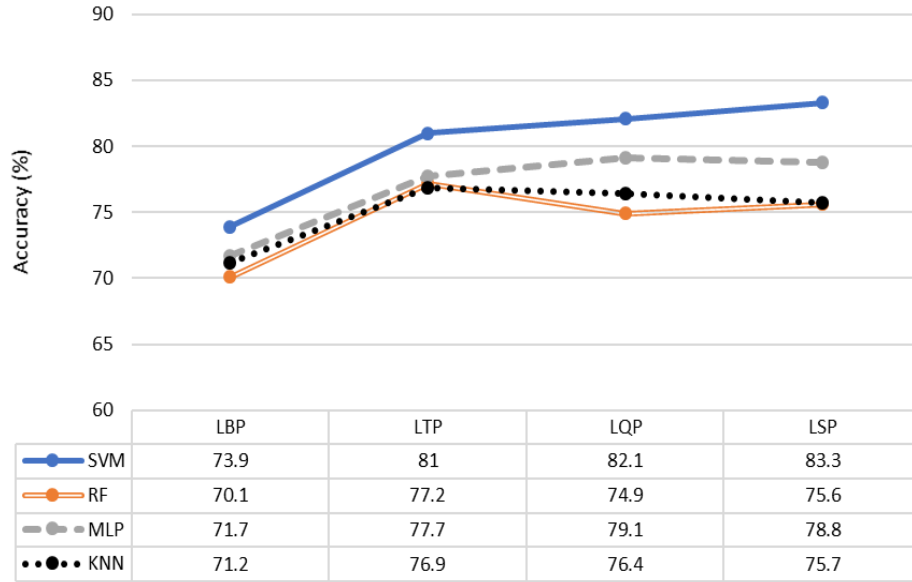


Figure 16: Comparison of maximum accuracies achieved by each operator across different classifiers when evaluated on the MIAS dataset.

For the InBreast dataset, Figures 18 and 19 show performance comparisons between all operators using different classifiers. Regarding maximum accuracy, the LSP operator achieved  $Acc = 80.5\%$  using the SVM classifier and outperformed the LBP, LTP and LQP operators by 6.7%, 6.4%, and 0.6%, respectively. An improvement also can be observed for the  $k$ -NN classifier as the LSP operator produced 79.7% compared to the other operators

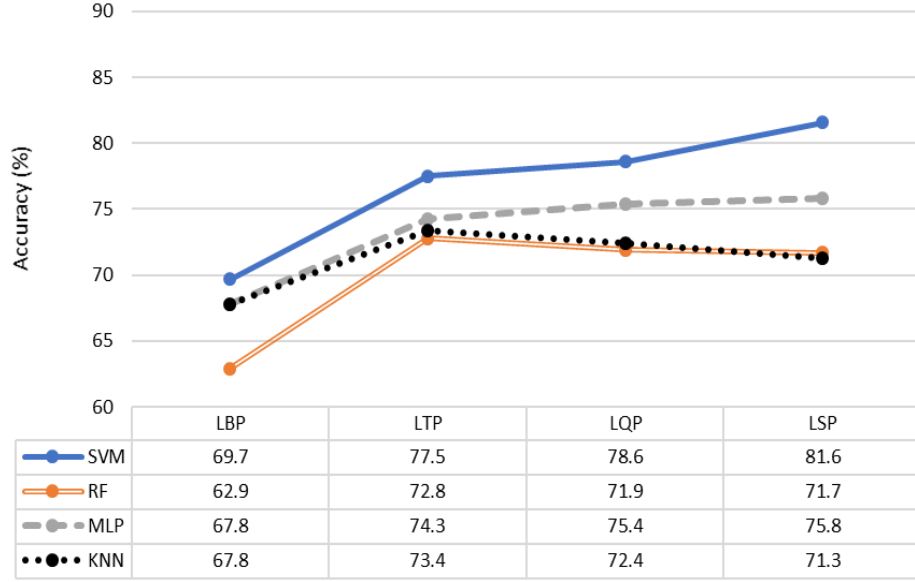


Figure 17: Comparison of average accuracies across different numbers of dominant patterns achieved by each operator across different classifiers when evaluated on the MIAS dataset.

532 which produced  $Acc < 75\%$ . For the RF classifier, the LSP operator pro-  
 533 duced 77.9%, at least 1.2% higher compared to the other operators. The  
 534 LQP operator produced a maximum accuracy of 80.1% which is 0.3% higher  
 535 than the LSP operator when using the MLP classifier. Once again it can  
 536 be observed that other operators always outperform the classification results  
 537 of the LBP operator regardless of classifiers. In terms of average accuracies,  
 538 the LSP operator once again outperforms the other operators when the SVM  
 539 classifier is employed with the best average accuracy of 77.1%. Similar trends  
 540 can be seen when using the  $k$ -NN classifier where the LSP features produced  
 541 the best average accuracy of 73.8% compared to 64.6%, 68.7% and 70.2%  
 542 produced by the LBP, LTP and LQP features, respectively. The LSP and  
 543 LTP operators produced similar results of 70.9% when the RF classifier is  
 544 employed which is 0.6% higher than the LQP operator. The LSP operator  
 545 produced the best average accuracy of 73.3% using the MLP classifier which  
 546 is slightly lower compared to LQP and LTP operators with 75.5% and 74.3%  
 547 accuracy, respectively. Once again experimental results suggest that the LSP  
 548 operator extracts more discriminant local features and performs better when  
 549 the SVM classifier is employed.

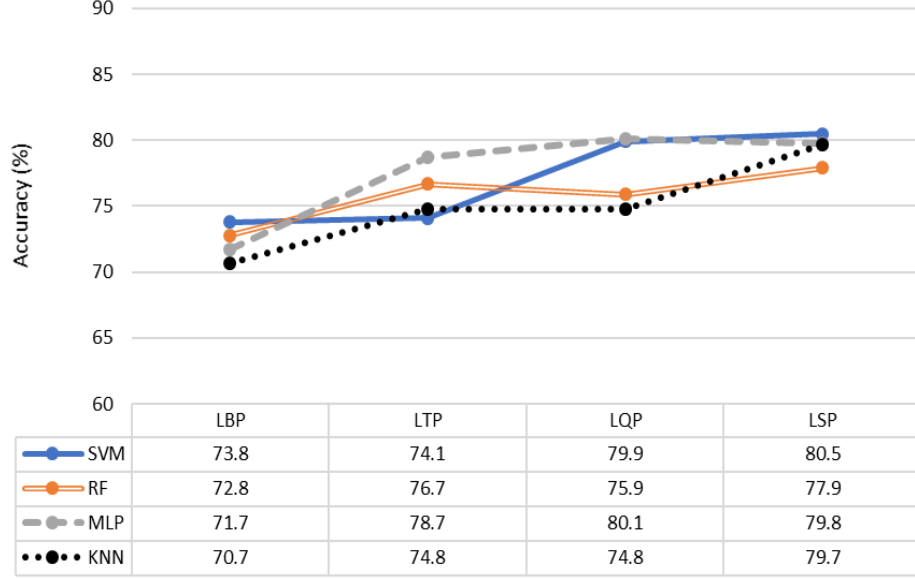


Figure 18: Comparison of maximum accuracies achieved by each operator across different classifiers when evaluated on the InBreast dataset.

## 7. Discussion

In this section we will discuss (i) the effects on accuracy when different numbers of dominant patterns are used in the classification, (ii) the effects of the radius ( $R$ ) and number of neighbours ( $P$ ) in the local patterns discriminatory levels; (iii) explanations on why multichannel local patterns produced more discriminant features in comparison to the original LBP operator, in (iv) extending to 9 and 11- encoding systems, (v) discussion of the existing studies in the literature, (vi) statistical analysis and (vii) future work.

### 7.1. Effects of Different Numbers of Dominant Patterns ( $n$ )

To investigate the effects of  $n$  on the classification accuracy for all operators described in this study we tested 60 different values from 90 to 99.9 (e.g.,  $n \in \{90, 90.1, 90.3, 90.5, \dots, 99.9\}$ ). In other words, we investigated the variation of classification accuracy by including local patterns which have a frequency of minimum 0.1% to 10% in the histogram feature (e.g. local patterns with a relative frequency of less than 0.1% to 10% are removed, resulting to a shorter histogram feature). Figure 20 shows the effects of  $n$

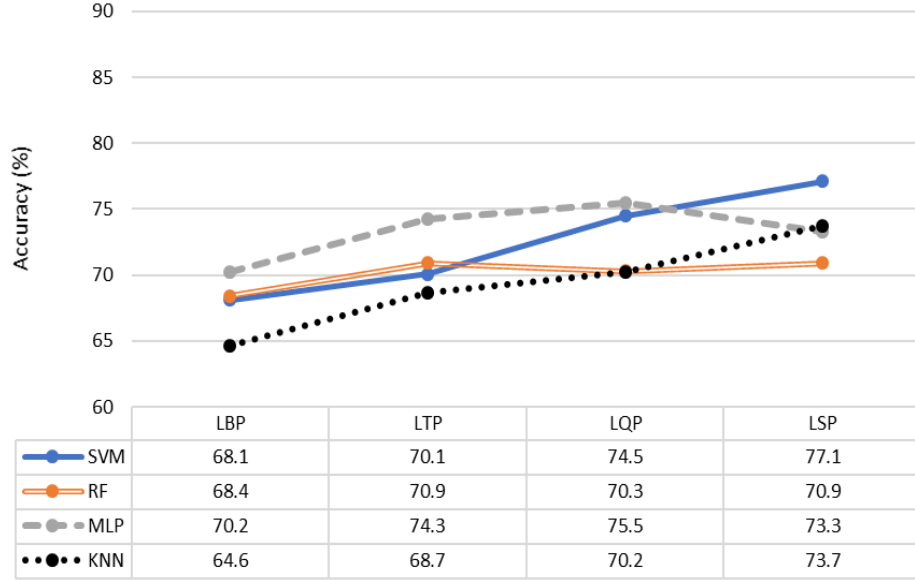


Figure 19: Comparison of average accuracies across different numbers of dominant patterns achieved by each operator across different classifiers when evaluated on the InBreast dataset.

on classification accuracy for LBP, LTP, LQP and LSP operators. In this experiment, we employed the SVM classifier as it produced better performance compared to the MLP, RF and  $k$ -NN classifiers. It can be observed that  $n$  plays an important role in getting the best classification accuracy. For the LBP operator it can be observed that a large variation of classification accuracy between 63% to 74% occurs, with a 9.71% standard deviation. The LQP operator produced a standard deviation of 9.25% which is higher than the LTP operator of 9.04%. The LSP operator produced a smaller standard deviation value of 9.04% (the same as LTP) which indicates that our proposed method does not only outperform the other operators on both datasets but also produced more consistent results.

## 7.2. Effects of Radius ( $R$ ) and Number of Neighbours ( $P$ )

In this section we are interested to know to what extent  $R$  and  $P$  affect the overall classification accuracy. For this purpose, we investigated seven different combination of  $R$  and  $P$  and tested each of them with all operators. Figures 21 and 22 show average maximum accuracies for both datasets using

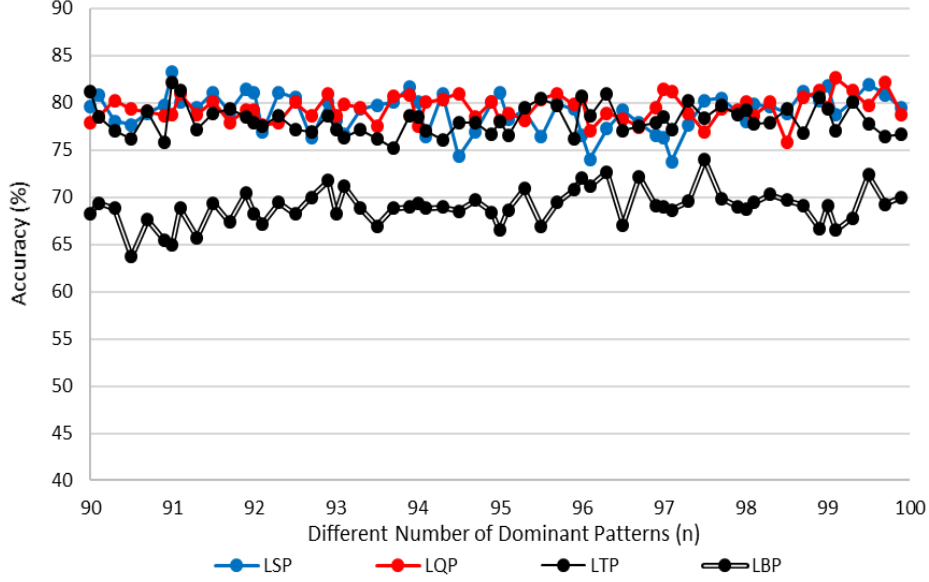


Figure 20: Effects of different numbers of dominant patterns ( $n$ ) across different operators ((P,R)=(18,9)) using SVM classifier on MIAS dataset.

the SVM classifier. For the LBP operator, local patterns extracted using smaller radii and less number of neighbours (e.g., (5, 10) and (5, 12)) tend to produce better results whereas the LTP operator produced its highest average maximum accuracy at (7, 14) ( $Acc > 78\%$ ). Similar results of 77.5% are obtained when the following parameters are employed: (7, 16), (9, 18) and (9, 20). However, when evaluated on the InBreast dataset, the highest average maximum accuracy (76%) is achieved at (9, 18) followed by parameters (7, 14) and (7, 16). For the LQP operator, the highest average maximum accuracy ( $Acc > 78\%$ ) is achieved at parameters (7, 14) followed by (9, 18) with only 0.5% difference. In the InBreast dataset, the LQP operator produced highest average maximum accuracy of 77.5% with parameters (9, 18). The LSP operator produced consistent average maximum accuracy when parameters (7, 14), (7, 16), (9, 18), (9, 20) and (11, 22) are used which is around 77%. Nevertheless, it produces at least 2.5% higher average maximum accuracy at (9, 20) compared to (7, 14), (7, 16), (9, 18) and (11, 22). These results indicate that when choosing the values for  $P$  and  $R$ , the following values (7, 16), (9, 18), (9, 20) are a good starting point. The LSP operator produced its best

599 results at (9, 20) using the SVM classification approach.

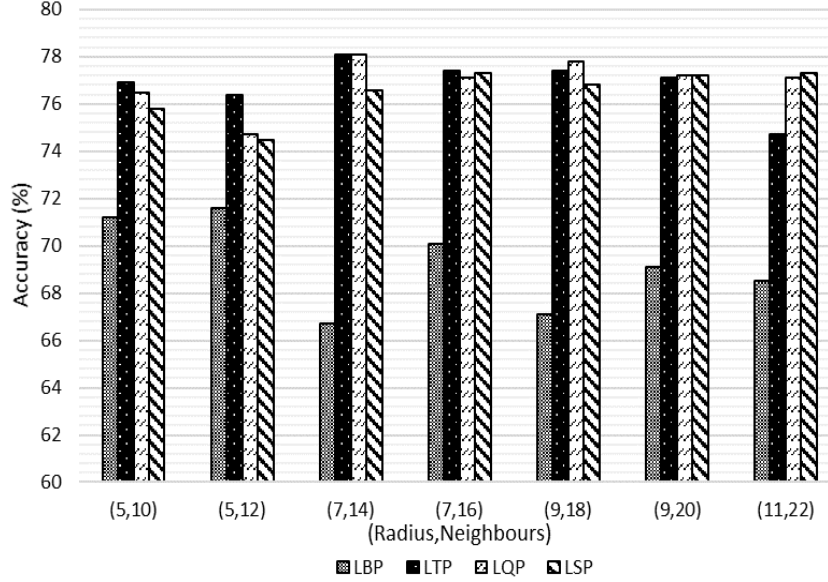


Figure 21: Average maximum accuracy (MIAS dataset) across different classifiers based on different operators and their associated parameters.

### 600 7.3. Why do multichannel local patterns work?

601 To answer this question we conducted several experiments by performing  
602 channel based classifications on the MIAS dataset. In the case of the  
603 LTP operator (three channels with two binary patterns), firstly we perform  
604 individual classification based on features extracted from the first binary  
605 patterns and secondly we concatenate features extracted from two or more  
606 channels. Table 3 shows our experimental results for the LTP, LQP and LSP  
607 operators. Note that the following parameters are used based on their best  
608 performance using the SVM classifier (see Figures 12, 13 and 14):  $LTP(7, 16)$   
609 with  $n = 93.5$ ,  $LBP(9, 18)$  with  $n = 99.9$  and  $LSP(9, 18)$  with  $n = 99.1$ .

610 It can be observed that classification based on features extracted from  
611 a single channel alone is insufficient. For the LTP operator, each chan-  
612 nel produced just over 73% but concatenating local patterns from all chan-  
613 nels produced 81%, yielding over 7% improvement. A similar case occurred  
614 for the LQP operator where binary patterns from a single channel pro-  
615 duced  $Acc < 70\%$ . However, combining binary patterns from the first three

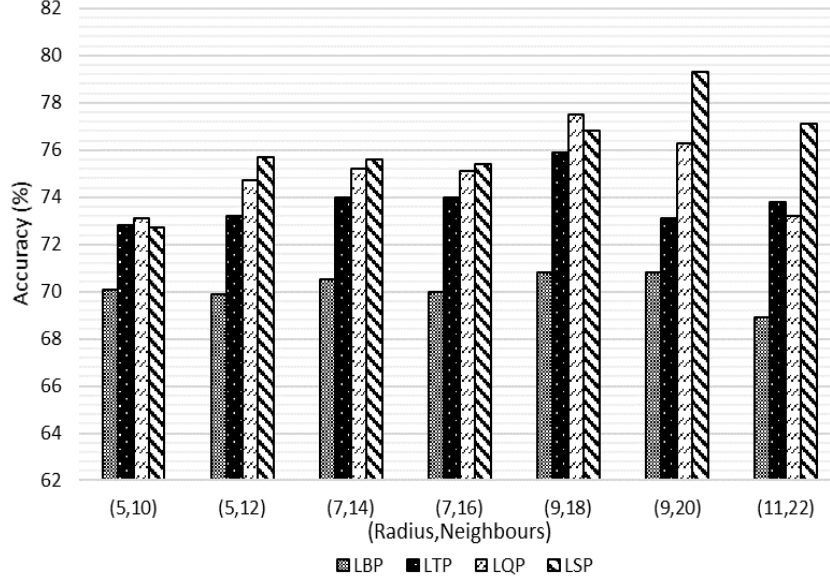


Figure 22: Average maximum accuracy (InBreast dataset) across different classifiers based on different operators and their associated parameters.

channels ( $Ch1 \oplus Ch2 \oplus Ch3$ ) yields a significant improvement 76.4% and concatenating all local patterns from all channels produced a maximum accuracy of 82.1%. For our proposed operator,  $Ch2$  and  $Ch5$  produced classification accuracy under 60% and  $Ch1$ ,  $Ch3$  and  $Ch4$  produced accuracy close to 70%. Concatenating all features from all channels resulted in  $Acc = 83.3\%$  however removing a less informative channel (e.g.,  $Ch2$ ) by considering  $Ch1 \oplus Ch3 \oplus Ch4 \oplus Ch5 \oplus Ch6$ , the classification accuracy improved to 83.8%.

Based on these experiments we found that encoding an image by dividing it into several channels captures more textural details which is difficult to capture with the original LBP operator. This can be clearly seen in Figures 8 and 9 where most accuracies achieved by the LBP operator are under 70% whereas the other operators consistently achieved above 75% accuracy. This also indicates that the LQP and LTP operators capture more details of the Corpus Mammar region in comparison to the LBP operator. Similarly, the LSP operator captures even more details compared to LQP and LTP operators. However, a higher order encoding system such as dividing into nine or eleven channels does not necessarily improve the classification accuracy

Operators	Channel ( $Ch$ )	Channels Concatenation ( $\oplus$ )
LTP	$Ch1 = 73.3 \pm 8.6$ $Ch2 = 74.1 \pm 9.1$	$Ch1 \oplus Ch2 = 81.95$
LQP	$Ch1 = 59.8 \pm 10.6$ $Ch2 = 65.9 \pm 9.6$ $Ch3 = 68.9 \pm 11.5$ $Ch4 = 66.5 \pm 9.9$	$Ch1 \oplus Ch2 = 69.3 \pm 8.3$ $Ch1 \oplus Ch2 \oplus Ch3 = 76.4 \pm 8.1$ $Ch1 \oplus Ch2 \oplus Ch3 \oplus Ch4 = 82.1 \pm 7.1$
LSP	$Ch1 = 68.9 \pm 9.4$ $Ch2 = 52.6 \pm 11.6$ $Ch3 = 70.1 \pm 8.8$ $Ch4 = 69.1 \pm 8.9$ $Ch5 = 59.3 \pm 11.3$ $Ch6 = 65.7 \pm 12.8$	$Ch1 \oplus Ch2 = 74.3 \pm 7.9$ $Ch1 \oplus Ch2 \oplus Ch3 = 75.1 \pm 7.9$ $Ch1 \oplus Ch2 \oplus Ch3 \oplus Ch4 = 70.1 \pm 10.1$ $Ch1 \oplus Ch2 \oplus Ch3 \oplus Ch4 \oplus Ch5 = 77.7 \pm 9.1$ $Ch1 \oplus Ch3 \oplus Ch4 \oplus Ch6 = 80.7 \pm 10.1$ $Ch1 \oplus Ch3 \oplus Ch4 \oplus Ch5 \oplus Ch6 = 83.8 \pm 9.8$ $Ch1 \oplus Ch2 \oplus Ch3 \oplus Ch4 \oplus Ch5 \oplus Ch6 = 83.3 \pm 8.8$

Table 3: Classification results (%) based on binary patterns extracted from individual channel (second column) and concatenated binary patterns from two or more channels

because it increases the number of features yielding more complex decision boundaries. An early indication can be observed based on our experiment results in Table 3 where best accuracy can be achieved by the LSP operator when excluding local patterns from  $Ch2$ . This reduces the number of features yielding a less complex decision boundaries in a feature space, hence boosting the overall performance of the method.

A channel encoding system also extracts a set of ‘weak’ features which produce satisfactory results on its own, but when combined they provide good results (similar to ensemble classifiers such as the Random Forests). For example, in Table 3 combining two or more local patterns from different channels always produce better results. This is due to each channel containing unique information which is not available in the other channels and they complement each other when combined. For example, individual classification in LQP produce under 70% accuracy but improve to over 76% and 82% when three and four channels are combined, respectively. Simi-



larly, all channels from the LSP operator are unable to provide sufficient information of the Corpus Mammæ region individually but when combining  $Ch1 \oplus Ch3 \oplus Ch4 \oplus Ch5 \oplus Ch6$  we achieved 83.8% accuracy.

Moreover, the threshold values which are determined automatically enable the operator to capture robust texture information as pixel information is encoded according to the local threshold values rather than a global threshold value as in the LTP and LQP operators. Local threshold values enable the LSP operator to capture textural information based on the topology and intensity distribution of the image and hence are more discriminant and tolerant to noise. In contrast, a global threshold value does not consider intensity distribution of a region. As a result, it may affect the discriminatory level of the local patterns.

#### 7.4. Extending to 9-encoding and 11-encoding systems

To further evaluate the performance of different encoding systems, we conducted several experiments by extending the LSP to 9- and 11-encoding systems both on the MIAS and InBreast datasets. Figures 23, 24, 25 and 26 show that the performance is decreasing as we use 9- and 11-encoding systems.

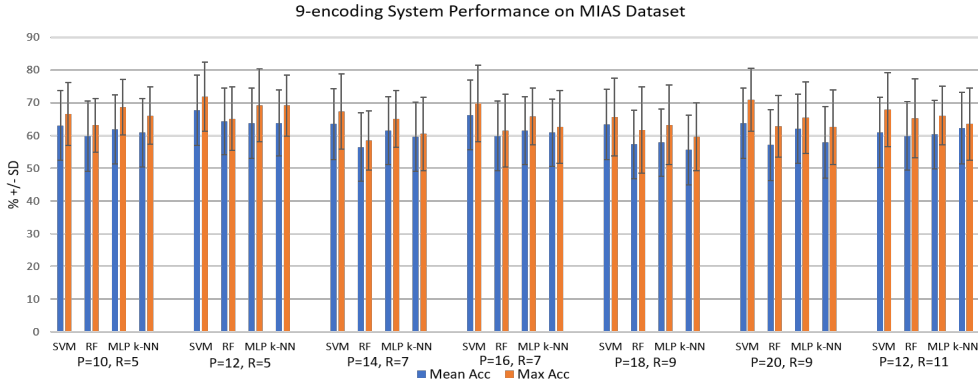


Figure 23: 9-encoding system performance on MIAS dataset

Experimental results suggest that the 7-encoding system the LSP achieved the best performance and eventually decreased to between 55% to 63%. From our own observation, the main reason for this is due to higher encoding systems (e.g. 9-encoding) producing very sparse features (containing many

671 zeros). This is similar when performing the 11-encoding system (in fact fea-  
 672 tures are even sparser). The main problem with sparse features is they are  
 673 most likely being ignored by most classifiers particularly tree-based machine  
 674 learning algorithms. Secondly, in most cases zeros are less informative mak-  
 675 ing the features less meaningful and hence less discriminant. As a result,  
 676 performance classification reduced significantly. In Figure 25 and 26 none of  
 677 the maximum accuracies achieved was above 70% whereas the LBP operators  
 678 achieved above this value in some cases.

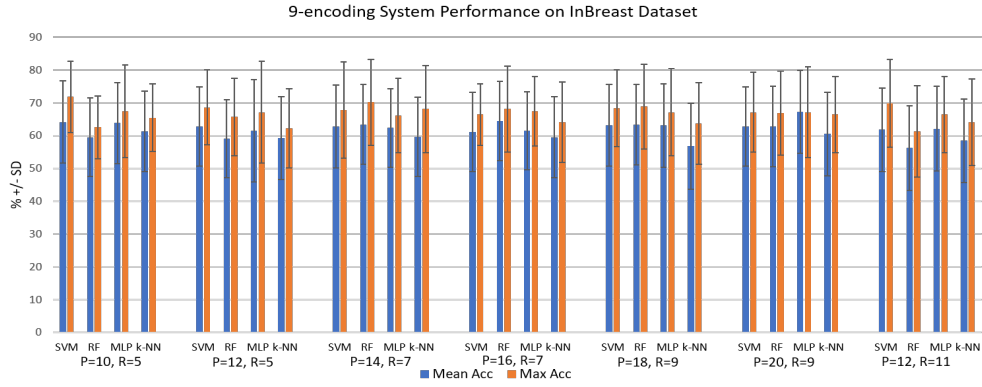


Figure 24: 9-encoding system performance on InBreast dataset

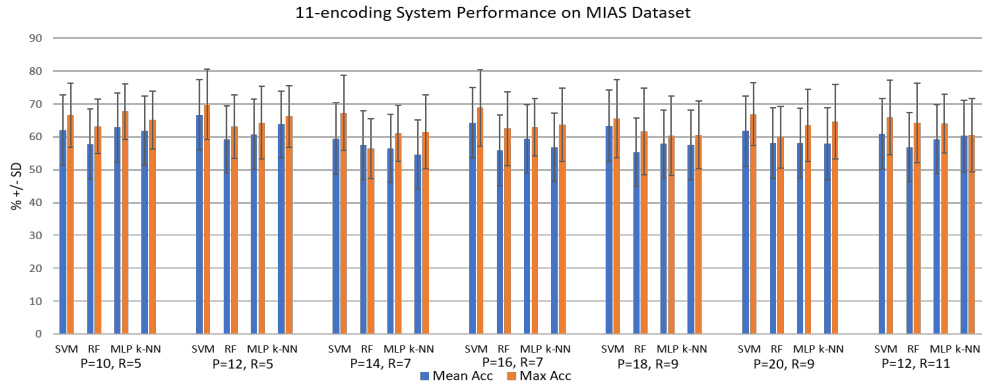


Figure 25: 11-encoding system performance on MIAS dataset

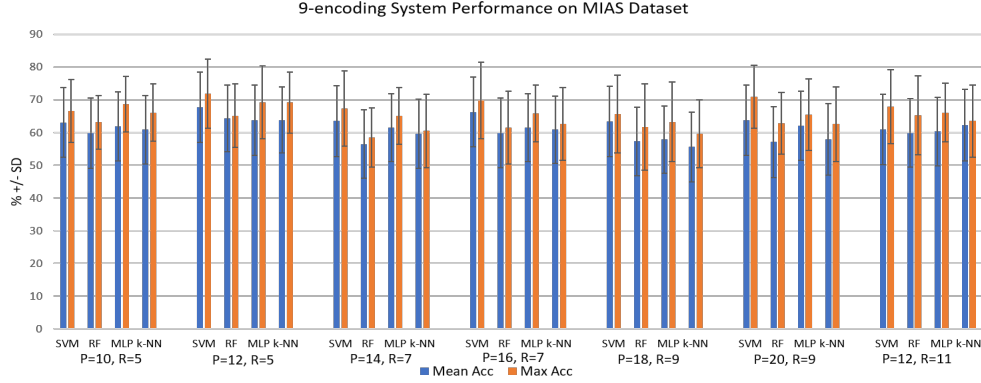


Figure 26: 11-encoding system performance on InBreast dataset

#### 679 7.5. Existing Methods in the Literature

680 The main goal of our study is to improve the performance of binary-  
681 based local pattern feature extraction methods, namely Local Binary Pat-  
682 tern (LBP), Local Ternary Pattern (LTP) and Local Quinary Pattern (LQP).  
683 Our study does not intend to quantitatively compare the performance of the  
684 LSP operator against other feature extraction methods. There are many fea-  
685 ture extraction methods in the literature, therefore quantitative comparison  
686 is extremely difficult. It should also be noted that quantitative compari-  
687 son with the other breast density methods in the literature is because most  
688 studies combine features from different feature extraction algorithms such as  
689 first-order statistical features, second-order statistical features, morphologi-  
690 cal features, features of Gabor filters, edge information, etc.

691 Many breast density classification methods have been developed over the  
692 last two decades. The best classification accuracy (based on BI-RADS classes  
693 (fourth edition)) reported in the literature is over 90% by the study by [14]  
694 followed by the study of [11] with 86% classification accuracy. However, these  
695 methods are computationally expensive due to different numbers of feature  
696 extraction techniques employed. Several techniques used to extract different  
697 types of features such as the first order statistical techniques to compute the  
698 local pixel intensity, the co-occurrence matrix to extract grey level distribu-  
699 tion, an adaptive thresholding technique to extract the region's morphological  
700 information, Gabor filter to obtain edges information, etc. Also, their meth-  
701 ods employed the Fuzzy C-Means clustering technique as a separate process  
702 to classify the breast tissue into different pixel intensities. Once again this

process is time consuming especially when dealing with large images such as mammograms. In contrast, the LSP operator (with accuracy close to 84%) extracts and processes all this information at once, hence reducing the computational complexity of the system.

Previously, [20] used the LTP operator to extract local information and achieved a promising result of 82.33% accuracy. However, this method suffers from having to deal with a large number of features (over one thousand) due to the multi-orientation approach (e.g., ten histograms from ten orientations concatenated). Our recent study, [21] reported the best accuracy of over 86% using multi-orientation LQP operator and combining other local patterns extracted using on ellipse neighbourhood. Nevertheless, this method contains several separate stages (e.g. extraction using different neighbourhood topologies (circle and ellipse), different orientations of resolutions which can be tedious whereas the LSP operator uses only one neighbourhood topology (e.g. a circle) with a single resolution. Other studies in the literature such as [11, 12, 13, 14, 15, 18, 19, 22, 23] reported accuracies under 80%. However, note that the purpose of our study is not to develop a breast density classification method that can necessarily outperform the other methods in the literature but to study the effects of various channel encoding techniques in the original LBP operator which have lead us to a more robust technique called the LSP operator, improving the performance of the LBP, LTP and LQP operators.

## 7.6. Statistical Analysis

We performed statistical analysis to investigate whether there is a significant difference/improvement at  $p \leq 0.05$  between the best results achieved by the LSP operator and the other operators discussed in this study on both datasets. For this purpose, we compared the best maximum (Max) and average (Avg) accuracies produced by the LSP with the best results of LBP, LTP and LQP operators. The  $p$ -value was computed using a  $t$ -test comparing each result of each operator with the best result achieved by the LSP operator according to their respective metric (e.g., Max *Acc* or Avg *Acc*). The size of population for each dataset are 103 (206 images) and 161 (322 images) for the InBreast and MIAS dataset, respectively.

Table 4 shows the  $p$ -values between the best accuracies achieved by the LSP operator and the other operators for both datasets. In terms of maximum accuracy, there is a significant improvement at  $p \leq 0.05$  when using the LSP features in classifying breast density on the MIAS ( $p = 0.0015$ ) and

Dataset	Results	LBP	LTP	LQP	LSP
MIAS	Max <i>Acc</i> (%)	73.8 $\pm$ 10.6	81 $\pm$ 9.5	82.1 $\pm$ 7.1	83.3 $\pm$ 8.8
	( <i>p</i> value)	(0.0015)	(0.2260)	(0.4019)	-
	Avg <i>Acc</i> (%)	69.7 $\pm$ 9.7	77.5 $\pm$ 9.2	78.6 $\pm$ 8.3	81.6 $\pm$ 9.1
	( <i>p</i> value)	(0.00001)	(0.0428)	(0.0433)	-
InBreast	Max <i>Acc</i> (%)	73.8 $\pm$ 9.9	78.7 $\pm$ 11.2	80.1 $\pm$ 10.5	80.5 $\pm$ 9.2
	( <i>p</i> value)	(0.0064)	(0.2646)	(0.4489)	-
	Avg <i>Acc</i> (%)	70.2 $\pm$ 11.7	74.3 $\pm$ 11.4	75.5 $\pm$ 11.3	77.1 $\pm$ 10.9
	( <i>p</i> value)	(0.0007)	(0.0137)	(0.2151)	-

Table 4: Classification results (%) and *p*-values between the best accuracies achieved by the LSP operator and the other operators discussed in this study.

740 InBreast ( $p = 0.0007$ ) datasets. Nevertheless, the best maximum results of  
741 the LSP operator are not statistically significant at  $p \leq 0.05$  in compari-  
742 son with the best accuracies achieved by the LTP and LQP operators. The  
743 *p*-values are 0.2260 and 0.4019 for the LTP and LQP operators, respectively  
744 when evaluated on the MIAS dataset which are similar when tested on the  
745 InBreast dataset,  $p = 0.2646$  and  $p = 0.4489$  for the LTP and LQP oper-  
746 ators, respectively. The results are statistically significant at  $p \leq 0.05$  on  
747 the MIAS dataset when comparing the best average accuracy produced by  
748 the LSP features ( $Acc = 81.6\%$ ) with the other features with *p*-values are  
749 0.00001, 0.0428 and 0.0433 for LBP, LTP and LQP respectively. Similarly,  
750 when tested on the InBreast dataset the best average accuracy produced by  
751 the LSP operator is significantly better in comparison to the results of LBP  
752 ( $p = 0.0007$ ) and LTP ( $p = 0.0137$ ) operators. However, the best average  
753 accuracy of LSP is not significant ( $p = 0.2151$ ) compared to the result of the  
754 LQP operator.

### 755 7.7. Limitations of our study

756 We highlight the following limitations of our study:

- 757 1. The sample dataset used to optimise the classifiers' parameters was  
758 taken from the MIAS dataset which only 8-bit contrast. This means  
759 the optimised parameters and testing results might be different when

testing with real clinical data (12 to 14 bits contrast). The reasons we chose the MIAS dataset rather than the InBreast dataset when optimising the classifiers' parameters are (a) it has larger number of images/cases hence giving us a larger number of test cases and (b) the number of cases for each class is more representative compared to the number of cases for each class in the InBreast dataset. For example, the number images of BI-RADS IV is 14 which is extremely small. Including cases in our dataset for parameter optimization, will make our testing dataset extremely imbalanced. Although in this study we used images from the MIAS dataset for parameter optimization, in a real clinical environment this procedure can be easily changed/adapted by using a sub-sample from the new dataset.

2. The MIAS dataset is a somewhat old database and hence does not represent the actual contrast representation of the latest mammograms. We included the MIAS dataset in our study due to (a) it is the most used and compared dataset in the literature because of its availability whereas other datasets are not easily accessible, (b) acquiring mammogram datasets is very difficult especially the ones with BI-RADS (version 4) classification and (c) other datasets do not classify each breast image based on BI-RADS density but are classified based on a small square patch based on tissue types such as 'fatty', 'glandular' or 'breast tissue'.
3. The parameters  $P$ ,  $R$  and  $n$  were tested/selected empirically. Nevertheless, our testing strategy is based on the previous studies of Ojala et al. [32], Tan and Triggs [33], Nanni et al. [34]. To the best of our knowledge this a common approach used in the literature especially for parameters  $P$  and  $R$ . For parameter  $n$ , we tested 60 different values from  $n = 90$  to  $n = 99.9$  with 0.2 interval (e.g. 90, 90.1, 90.3, 90.5, etc). This should be sufficient to investigate the performance variations across different  $n$  values.
4. The number of cases used in the development of parameter optimization is small. However, we prefer to optimise the number of images for testing purposes and therefore, we limit the number of cases for parameter optimisation. It is more important to test the model performance on a larger size of dataset rather than on a smaller dataset.

#### 7.8. Future Work

For future work we plan to consider the following:

1. Since multichannel based operators such as LTP, LQP and LSP require the number of dominant patterns ( $n$ ) to obtain an optimum classification accuracy, we plan to develop a statistical method that can automatically determine an optimum value of  $n$ .
2. We also interested to investigate the performance of LBP's two-channel variants such as M-ELBP, ELBP, MRELBP, STLBP and CLBP when multichannel based operators are being applied.
3. As indicated in Table 3, combining all local patterns from all channels does not necessarily produce optimal accuracy. For example, the LSP operator achieved its highest accuracy when excluding *Ch2*. Therefore, we plan to develop a method that can automatically determine the most informative local patterns from each channel and in the classification phase only combine the most discriminant features. This can be done by measuring mutual information among the histogram features and combining the ones with the least overlapping information.
4. We plan to develop a deep learning based network to extract local features of the Corpus Mammæ region and combine them with features extracted with the LQP or LSP operators.

## 8. Summary and Conclusions

In this paper, we studied the effects of various channel encoding techniques in local pattern extraction which have led us to a more robust technique called the LSP operator. We investigated the following aspects in our study:

1. We made comparisons between the LBP, LTP, LQP and LSP operators based on MIAS and InBreast datasets. Experimental results suggest that the proposed texture operator outperformed the other operators on both datasets.
2. We investigated whether the choice of a classifier can influence the performance of the system. We employed four different classifiers and found that in many cases the SVM classifier outperforms the other classifiers. Local patterns extracted with the LTP operator tend to produce better results when the MLP or  $k$ -NN classifier is employed whereas the SVM classifier produced better accuracy using LQP and LSP features.

- 831 3. We conducted experiments to investigate whether the operators dis-  
832 cussed in this study have a particular preference regarding  $R$  and  $P$ .  
833 We found that in many cases LBP features produced similar results of  
834 under 70% accuracy. However, the LSP features tend to produce better  
835 classification results when  $LTP(7, 14)$  and  $LTP(9, 18)$  are used. On the  
836 other hand, both LQP and LSP operators tend to produce consistent  
837 results when  $R \in \{7, 9\}$  and  $P \in \{16, 18, 20\}$  are used.
- 838 4. We studied the effects of selecting different numbers of dominant pat-  
839 terns and experimental results show that the value of  $n$  plays an impor-  
840 tant role in obtaining optimal accuracy. The LSP and LTP operators  
841 produced variation accuracy of 9.04% compared to 9.25% and 9.71%  
842 for LQP and LBP, respectively.
- 843 5. We also studied channel based classification and concatenating local  
844 patterns from different channels. Experimental results suggest that  
845 each channel contains unique features which are not available in the  
846 other channels and combining these features yielded to better classifi-  
847 cation results. We also found that removing the most uninformative  
848 channel in LSP features improves the classification accuracy by 0.5%.
- 849 6. We developed simple methods to automatically determine the threshold  
850 values in LSP by computing the 25<sup>th</sup> percentile and 75<sup>th</sup> percentile of  
851 the neighbouring pixels.

852 In conclusion, we have studied various channel encoding techniques in  
853 LBP, LTP and LQP operators where we found that multichannel local pat-  
854 terns are more robust in discriminating different classes of breast density.  
855 This study has led us to the development of a seven-encoding technique call  
856 LSP operator to capture more texture details within the Corpus Mammae re-  
857 gion. We also introduced a simple method to automatically determine the  
858 threshold values in the LSP operator by computing the first order statisti-  
859 cal values of the neighbourhood pixels. Experimental results show that the  
860 LSP features outperformed the LBP, LTP and LQP operators on both MIAS  
861 and InBreast datasets. The LSP features produced a maximum accuracy of  
862 83.8% using the SVM classifier when combining local patterns from  $Ch1$ ,  
863  $Ch3$ ,  $Ch4$ ,  $Ch5$  and  $Ch6$  on the MIAS dataset. The LBP, LTP and LQP  
864 features achieved maximum accuracies of 73.8%, 81% and 82.1%, respec-  
865 tively. Based on these experimental results, the proposed seven-encoding  
866 system approach (LSP) is shown to be a robust and more consistent texture  
867 operator.



## 868 Declaration of competing interest

869 None Declared.

## 870 Acknowledgment

871 This research was undertaken as part of the Decision Support and In-  
872 formation Management System for Breast Cancer (DESIREE) project. The  
873 project has received funding from the European Union's Horizon 2020 re-  
874 search and innovation programme under grant agreement No 690238.

## 875 References

- 876 [1] American Cancer Society, Breast cancer facts and figures  
877 2017/2018, 2018. Available at [https://www.cancer.org/content/  
878 dam/cancer-org/research/cancer-facts-and-statistics/  
879 breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.  
880 pdf](https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf) Accessed: 2017-06-23.
- 881 [2] C. E. DeSantis, J. Ma, A. G. Sauer, L. A. Newman, A. Jemal, Breast  
882 cancer statistics, 2017, racial disparity in mortality by state., CA: A  
883 Cancer Journal for Clinicians 67 (2017) 439–448.
- 884 [3] Breast Cancer, UK, Breast cancer care. facts and statistics 2018,  
885 2018. Available at [https://www.breastcancercare.org.uk/  
886 about-us/media/press-pack-breast-cancer-awareness-month/  
887 facts-statistics](https://www.breastcancercare.org.uk/about-us/media/press-pack-breast-cancer-awareness-month/facts-statistics) Accessed: 2017-05-23.
- 888 [4] Breast Cancer, UK, Breast cancer symptoms, 2018. Avail-  
889 able at [https://www.cancerresearchuk.org/about-cancer/  
890 breast-cancer/symptoms](https://www.cancerresearchuk.org/about-cancer/breast-cancer/symptoms) Accessed: 2017-06-26.
- 891 [5] Breast Cancer, U.S. breast cancer statistics, 2016. Available at [http:  
892 //www.breastcancer.org/symptoms/understand\\_bc/statistics](http://www.breastcancer.org/symptoms/understand_bc/statistics)  
893 Accessed: 2017-06-28.
- 894 [6] Breastcancer.org, Dense breast, 2018. Available at [https://www.  
895 breastcancer.org/risk/factors/dense\\_breasts](https://www.breastcancer.org/risk/factors/dense_breasts) Accessed: 2017-06-  
896 28.

- 897 [7] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi,  
898 M. Ghafoorian, J. A. van der Laak, B. van Ginneken, C. I. Sánchez,  
899 A survey on deep learning in medical image analysis, *Med. Image Anal.*  
900 42 (2017) 60–80.
- 901 [8] A. Hamidinekoo, E. Denton, A. Rampun, K. Honnor, R. Zwiggelaar,  
902 Deep learning in mammography and breast histology, an overview and  
903 future trends, *Med Image Anal* 47 (2018) 45–67.
- 904 [9] N. F. Boyd, J. W. Byng, R. A. Jong, E. K. Fishell, L. E. Little, A. B.  
905 Miller, G. A. Lockwood, D. L. Tritchler, M. J. Yaffe, Quantitative  
906 classification of mammographic densities and breast cancer risk: Results  
907 from the canadian national breast screening study, *J. Nat. Cancer Inst.*  
908 87 (1995) 670–675.
- 909 [10] J. Suckling et al., The mammographic image analysis society digital  
910 mammogram database, in: *Proc. Excerpta Med. Int. Congr. Ser.*, pp.  
911 375–378.
- 912 [11] A. Oliver, J. Freixenet, R. Martí, E. P. J. Pont, E. R. E. Denton,  
913 R. Zwiggelaar, A novel breast tissue density classification methodol-  
914 ogy, *IEEE Trans. Inf. Technol. Biomed.* 12 (2008) 55–65.
- 915 [12] K. Bovis, S. Singh, Classification of mammographic breast density using  
916 a combined classifier paradigm, in: *Proc. Int. Workshop on Digital*  
917 *Mammography, Lecture Notes in Computer Science*, Springer, 2002, pp.  
918 177–180.
- 919 [13] M. Muštra, M. Grgić, K. Delač, Breast density classification using mul-  
920 tiple feature selection, *Automatika* 53 (2012) 362–372.
- 921 [14] N. M. Parthaláin, R. Jensen, Q. Shen, R. Zwiggelaar, Fuzzy-rough  
922 approaches for mammographic risk analysis, *Intell. Data Anal.* 14 (2010)  
923 225–244.
- 924 [15] D. Raba, J. Martí, R. Martí, M. Peracaula, Breast mammography asym-  
925 metry estimation based on fractal and texture analysis, in: *Proc. Com-*  
926 *puted Aided Radiology and Surgery*, Berlin, Germany, volume 1398.

- 927 [16] H. Li, M. L. Giger, O. I. Olopade, A. Margolis, L. Lan, M. R. Chinander,  
928 Computerized texture analysis of mammographic parenchymal patterns  
929 of digitized mammograms, *Acad. Radiol.* 12 (2005) 863–73.
- 930 [17] A. Bosch, X. Munoz, A. Oliver, J. Martí, Modeling and classifying  
931 breast tissue density in mammograms, in: *Proc IEEE Comput. Soc.  
932 Conf. Comput. Vis. Pattern. Recognit. (CVPR)*, volume 2, pp. 1552–  
933 1558.
- 934 [18] Z. Chen, E. Denton, R. Zwigelaar, Local feature based mammographic  
935 tissue pattern modelling and breast density classification, in: *Proc.  
936 4th International Conference on Biomedical Engineering and Informatics  
937 (BMEI)*, volume 1, pp. 351–355.
- 938 [19] S. Petroudi, T. Kadir, M. Brady, Automatic classification of mammo-  
939 graphic parenchymal patterns: A statistical approach, in: *Proc. IEEE  
940 Conf. Eng. Med. Biol. Soc.*, volume 1, pp. 798–801.
- 941 [20] A. Rampun, P. J. Morrow, B. W. Scotney, R. J. Winder, Breast density  
942 classification using local ternary patterns in mammograms, in: *Proc. Im-  
943 age Analysis and Recognition*, Springer International Publishing, Cham,  
944 2017, pp. 463–470.
- 945 [21] A. Rampun, B. W. Scotney, P. J. Morrow, H. Wang, J. Winder, Breast  
946 density classification using local quinary patterns with various neigh-  
947 bourhood topologies, *J. Imaging* 4 (2018) 14.
- 948 [22] M. George, A. Rampun, E. Denton, R. Zwigelaar, Mammographic  
949 ellipse modelling towards birads density classification, in: *Proc. Inter-  
950 national Workshop on Breast Imaging (IWDM) 2016: Breast Imaging*,  
951 pp. 423–430.
- 952 [23] M. George, R. Zwigelaar, Breast tissue classification using local binary  
953 pattern variants: A comparative study (in press), in: *Proc. International  
954 Workshop on Breast Imaging (IWDM) 2016: Breast Imaging*.
- 955 [24] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Méridaa, C. I.  
956 Sánchez, R. Manna, A. den Heeten, N. Karssemeijer, Large scale deep  
957 learning for computer aided detection of mammographic lesions, *Med.  
958 Image Anal.* 35 (2017) 303–312.

- 959 [25] A. Rampun, K. López-Linares, P. J. Morrow, B. W. Scotney, H. Wang,  
960 I. G. Ocaña, G. Maclair, R. Zwigelaar, M. A. González Ballester,  
961 I. Macía, Breast pectoral muscle segmentation in mammograms using  
962 a modified holistically-nested edge detection network, *Medical Image*  
963 *Analysis* 57 (2019) 1–17.
- 964 [26] A. A. Mohamed, W. A. Berg, H. Peng, Y. Luo, R. C. Jankowitz, S. Wu.,  
965 A deep learning method for classifying mammographic breast density  
966 categories, *Med. Phys.* 45 (2017) 314–321.
- 967 [27] S. Li, J. Wei, C. Heang-Ping, M. A. Helvie, M. A. Roubidoux, Y. Lu,  
968 C. Zhou, L. M. Hadjiiski, R. K. Samala, Computer-aided assessment  
969 of breast density: comparison of supervised deep learning and feature-  
970 based statistical learning, *Phys. Med. Biol.* 63 (2018) 15.
- 971 [28] M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel,  
972 C. M. Vachon, K. Holland, R. R. Winkel, N. Karssemeijer, M. Lill-  
973 holm, Unsupervised deep learning applied to breast density segmenta-  
974 tion and mammographic risk scoring, *IEEE Trans. Med. Imaging.* 35  
975 (2016) 1322–1331.
- 976 [29] J. Lee, R. M. Nishikawa, Automated mammographic breast density  
977 estimation using a fully convolutional network, *Med. Phys.* 45 (2018)  
978 1178–1190.
- 979 [30] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, J. S.  
980 Cardoso, Inbreast: toward a full-field digital mammographic database,  
981 *Acad. Radiol.* 19 (2011) 236–428.
- 982 [31] A. Rampun, P. J. Morrow, B. W. Scotney, R. J. Winder, Fully auto-  
983 mated breast boundary and pectoral muscle segmentation in mammo-  
984 grams, *Artif. Intell. Med.* 79 (2017) 28–41.
- 985 [32] T. Ojala, M. Pietikainen, T. Maenpaar, Multiresolution gray-scale and  
986 rotation invariant texture classification with local binary patterns, *IEEE*  
987 *Trans. Pattern Anal. Mach. Intell.* 24 (2002) 971–987.
- 988 [33] X. Tan, B. Triggs, Enhanced local texture feature sets for face recogni-  
989 tion under difficult lighting conditions, *IEEE Trans. Image Process.* 19  
990 (2010) 1635–1650.

- 991 [34] L. Nanni, A. Luminia, S. Brahnem, Local binary patterns variants as  
992 texture descriptors for medical image analysis, *Artif. Intell. Med.* 49  
993 (2010) 117–125.
- 994 [35] L. Liu, L. Zhao, Y. Long, G. Kuang, P. Fieguth, Extended local binary  
995 patterns for texture classification, *Image Vision Comput.* 30 (2012) 86  
996 – 99.
- 997 [36] S. Liao, A. C. S. Chung, Face recognition by using enlongated local  
998 binary patterns with average maximum distance gradient magnitude,  
999 in: *Proc. Computer Vision—ACCV 2007*, volume 4844, Lecture Notes  
1000 in Computer Science, 2007, pp. 672–679.
- 1001 [37] L. Liu, S. Lao, P. W. Fieguth, Y. Guo, X. Wang, M. Pietikäinen, Median  
1002 robust extended local binary pattern for texture classification, *IEEE*  
1003 *Trans. Image Process.* 25 (2016) 1368–1381.
- 1004 [38] J. Yang, S. Wang, Z. Lei, Y. Zhao, S. Z. Li, Spatio-temporal lbp based  
1005 moving object segmentation in compressed domain, in: *Proc. IEEE*  
1006 *Ninth International Conference on Advanced Video and Signal-Based*  
1007 *Surveillance*, pp. 252–257.
- 1008 [39] Z. Guo, L. Zhang, D. Zhang, A completed modeling of local binary  
1009 pattern operator for texture classification, *IEEE Trans. Image Process.*  
1010 19 (2010) 1657–1663.
- 1011 [40] Y. Gio, G. Zhao, M. Pietikäinen, Discriminative features for feature  
1012 description, *Pattern Recognit.* 45 (2012) 3834–3843.
- 1013 [41] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Wit-  
1014 ten, The weka data mining software: an update, *ACM SIGKDD explo-*  
1015 *rations newsletter* 11 (2009) 10–18.

# Breast Density Classification in Mammograms: An Investigation of Encoding Techniques in Binary-based Local Patterns

Andrik Rampun<sup>a,b,\*</sup>, Philip J. Morrow<sup>b</sup>, Bryan W. Scotney<sup>b</sup>, Hui Wang<sup>c</sup>

<sup>a</sup>*Academic Unit of Radiology, Department of Infection, Immunity and cardiovascular Disease, Sheffield University, S10 2RX, UK*

<sup>b</sup>*School of Computing, Ulster University, Jordanstown, Northern Ireland, BT37 0QB, UK*

<sup>c</sup>*School of Computing, Ulster University, Jordanstown, Newtownabbey, Northern Ireland, BT37 0QB, UK*

---

## Abstract

We investigate various channel encoding techniques when applied to breast density classification in mammograms specifically local binary, ternary and quinary encoding approaches are considered. Subsequently, we propose a new encoding approach based on a seven-encoding technique yielding a new local pattern operator called a local septenary pattern operator. Experimental results suggest that the proposed local pattern operator is robust and outperforms the other encoding techniques when evaluated on the Mammographic Image Analysis Society (MIAS) and InBreast datasets. The local septenary patterns achieved a maximum classification accuracy of 83.3% and 80.5% on the MIAS and InBreast datasets, respectively. The closest comparison achieved by the other local pattern operators is the local quinary operator producing 82.1% (MIAS) and 80.1% (InBreast) maximum accuracies, respectively.

*Keywords:* Breast mammography, Breast Density, Local Binary Patterns, Local Ternary Patterns, Local Quinary Patterns, Local Septenary Patterns

---

---

\*Corresponding author

*Email address:* [y.rampun@sheffield.ac.uk](mailto:y.rampun@sheffield.ac.uk) (Andrik Rampun)

## 1. Introduction

Breast cancer is the leading cause of death in women. Although younger women can also get breast cancer, women over 50 years old have a much higher chance of being affected contributing to over 90% of the cases [1]. In 2017, according to the recent breast cancer statistics report by [2], more than 250,000 cases of breast cancer were diagnosed in the United States. Breast Cancer UK revealed that over 11,000 women died in the United Kingdom (UK) in 2016 [3]. Although the causes are not fully understood, several risk factors such as age, family history, overweight/obesity and excessive use of alcohol were found to be the risk contributors of the disease. According to the National Health Services (NHS) in the UK, there are several symptoms of breast cancer such as the appearance of lesions, micro-calcification, breast density and change of size and shape of the breast [4]. Nowadays, mammography is the most common imaging technology used for screening breast cancer to find early symptoms of breast cancer. Other alternative technologies are Ultrasound and Magnetic Resonance Imaging (MRI). In current clinical practice, radiologists have to examine each mammogram of a patient to find any signs of abnormality and doctors can tell whether a tumour is benign or malignant through biopsy tests. Although the overall current clinical methods have significantly improved in the last two decades, there are still some deficiencies such as inconsistency among radiologists, and the fact that it is time-consuming and invasive.

Women with a dense breast are six times more likely to develop cancer mainly because a tumour has a similar appearance to dense tissue, making it difficult to detect in mammograms [5, 6]. Several studies have indicated that there is a significant correlation between breast density and developing breast cancer [5, 6]. In most cases, dense breasts contain more glandular and fibrous tissue. In mammograms, breast regions with dense tissue tend to have more brighter pixels with fatty tissue usually represented by darker pixels. During the screening procedure, breast density estimation is visually assessed by radiologists and classified into four groups based on the amount of dense tissue. The Breast Imaging-Reporting and Data System (BI-RADS) fourth edition is one of the guidelines used in breast density assessment and contains four main categories:

1. BI-RADS I (0-25% dense tissues and mostly fatty)
2. BI-RADS II (26-50% dense tissues with some scattered density)

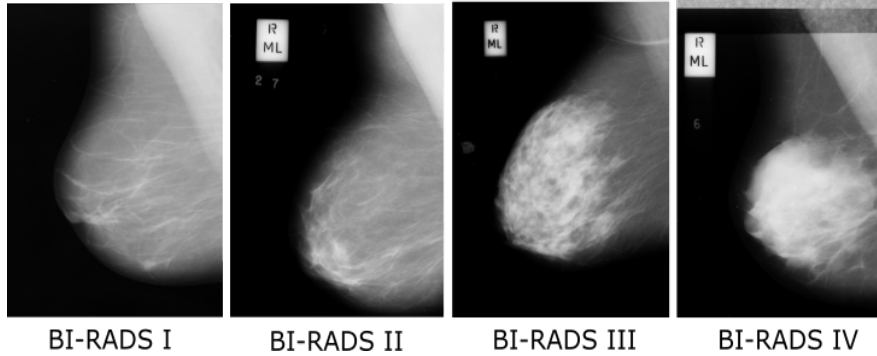


Figure 1: An illustration of breast density according to the BI-RADS guideline fourth edition taken from the the Mammographic Image Analysis Society (MIAS) dataset.

- 37 3. BI-RADS III (51-75% dense tissues and many areas of fibrous and
- 38 glandular tissue that are evenly distributed)
- 39 4. BI-RADS IV (over 75% dense tissues and breasts have a lot of fibrous
- 40 and glandular tissue which makes it hard to see a cancer because it can
- 41 blend in with the normal tissue)

42 Although manual assessment can be done by radiologists, this task is

43 time-consuming and often suffers from variability between radiologists. For

44 example, radiologists with more experience tend to produce more consis-

45 tent results compared to less experienced radiologists. Moreover, manual

46 assessment also could lead to false negatives/positives due to fatigue during

47 diagnostic decision making which could influence the final outcome. Figure

48 1 shows examples of four breasts with their associated BI-RADS classes. In

49 this study we use the fourth edition for simplicity because all datasets are

50 annotated for ground truth based on the BI-RADS fourth edition. However,

51 we are aware that the most recent BI-RADS guidelines is the fifth edition.

## 52 2. Computer-aided diagnosis (CAD) systems

53 Computer-aided diagnosis (CAD) systems can assist clinicians regard-

54 ing efficiency, effectiveness and consistency. CAD systems can assess lesions

55 un-invasively and make predictions as to whether a lesion is benign or malig-

56 nant. Furthermore, CAD can be used as a ‘second reader’ to assist clinicians

57 in diagnostic decision making particularly in cases where doctors are not

58 certain about their decision. However, developing a CAD system that can



replicate radiologists’ knowledge requires a significant amount of time and effort. Machine learning is a sub-component of Artificial Intelligence which enables machines to learn and predict a future occurrence of an event. With the use of machine learning and image processing techniques, it is possible to ‘teach’ computers to learn the appearance/characteristics of breast regions during the training phase and subsequently make a prediction based on the information/knowledge learned.

In breast CAD systems, one of the ways to characterise a breast region in a mammogram is by studying its texture appearance in order to extract meaningful information. In the last decades, many CAD systems for breast imaging have been developed and the majority of them used texture information to investigate the appearance of abnormality such as lesions, dense tissue and micro-calcification clusters. Texture information can be extracted using different mathematical operations such as filter-based, histograms, grey-level distributions, or statistical and probability techniques. Since 2012, the success of deep learning in image classification and segmentation has been overwhelming computer scientists. As a result, hundreds of papers about deep learning for medical image analysis have been published according to the studies by [7] and [8]. Nevertheless, the primary deficiency of deep learning based approaches is the need for a large number of datasets with breast density annotation which can be very challenging in the medical domain. To the best of our knowledge, none of the studies in the literature has attempted using deep learning for four-class breast density classification. The majority of them performed either two-class (low risk (BI-RADS I and II) versus high risk (BI-RADS III and IV)) or three-class (fatty, glandular and dense) classification. On the other hand, many studies have attempted four-class classification using conventional machine learning in conjunction with texture information.

In this study, we are interested in investigating the effects of various encoding techniques in local pattern extraction by dividing the binary patterns into two (local binary patterns (LBP)), three (local ternary patterns (LTP)) and five (local quinary patterns (LQP)) channels. Subsequently, we propose a local septenary patterns (LSP) operator which divides binary patterns into seven channels. Furthermore, we also study the discriminatory level of local patterns extracted from each channel which can help us to determine more informative texture descriptors. The paper has the following contributions:

1. To the best of our knowledge this is the first study attempting to inves-

- 96        tigate the effects of different encoding techniques covering LBP, LTP  
 97        and LQP applied to breast density classification.
- 98        2. We introduce a variant of LBP (two-encoding technique), LTP (three-  
 99        encoding technique) and LQP (five-encoding technique) called local  
 100        septenary patterns (LSP).
  - 101        3. We introduce several ways of improving local patterns' discriminatory  
 102        level in the application to breast density classification via a multichan-  
 103        nel concatenation approach that enables us to combine local binary  
 104        patterns from different channels.
  - 105        4. Whilst threshold values in LTP and LQP need to be set manually by a  
 106        user, we introduce a simple adaptive way to determine threshold values  
 107        in LSP based on the intensity distribution of the neighbourhood.
  - 108        5. Finally we also make a quantitative assessment on each of the channels  
 109        which can reveal more informative texture descriptors.

110        Note that our study does not attempt to develop a new CAD system for  
 111        breast density classification but to investigate the effects on discriminatory  
 112        levels of local patterns using various encoding techniques.

### 113    **3. Breast Anatomy in a Mammogram**

114        Figure 2 shows breast anatomy in a mediolateral-oblique mammogram  
 115        (MLO). In many cases, the upper retroglandular region contains fatty tissue  
 116        and the Corpus Mammarum region contains glandular and dense tissue (if the  
 117        breast is categorised BI-RADS III or IV). Note that usually dense tissue has  
 118        brighter pixels compared to glandular tissue and these tissues have a similar  
 119        appearance to the pectoral muscle region. Since many biological activities in  
 120        the Corpus Mammarum region are due to more Lobules and Ducts, dense tissue  
 121        mostly starts its development in this area.

### 122    **4. Literature Review**

123        Breast density classification in mammograms is one of the most popu-  
 124        lar topics in breast CAD systems together with mass and micro-calcification  
 125        clusters classifications. In a majority of the cases, textures have been a pop-  
 126        ular choice for most authors mainly due to its efficiency and effectiveness.  
 127        Previously, manual and adaptive thresholding techniques have been used to

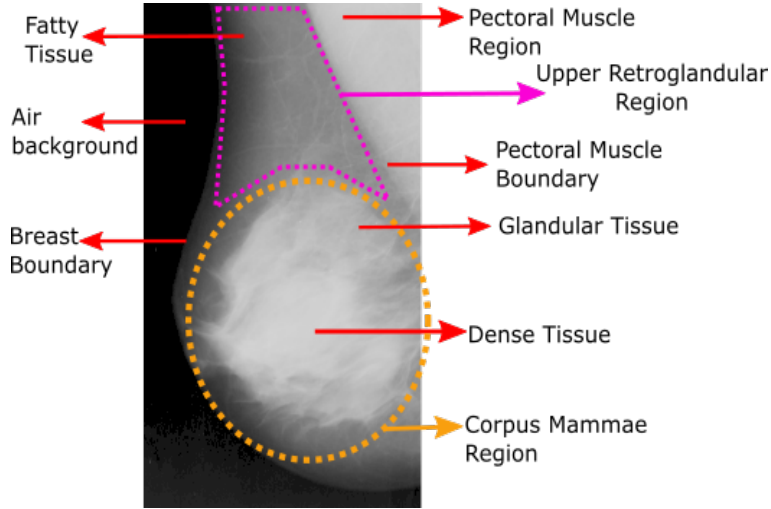


Figure 2: An overview of breast anatomy in a mediolateral-oblique mammogram (MLO). Note that the contrast of the image has been adjusted from the original image for improve visualisation of the breast boundary.

128 estimate the amount of dense tissue. For example, [9] developed an inter-  
 129 active thresholding software called Cumulus to segment regions with dense  
 130 tissue by manually tuning a grey-level threshold value. Due to the difficulty  
 131 in getting annotated mammograms, this topic did not get much attention  
 132 from computer scientists until the MIAS dataset was made available publicly  
 133 in 1994 [10].

134 The first- and second-order (e.g., Grey Level Co-occurrence Matrix (GLCM))  
 135 statistical features are among the most popular texture descriptors. For ex-  
 136 ample the studies of [11], [12], [13] and [14] all used these texture descriptors  
 137 by considering different orientations and resolutions. The methods of [11]  
 138 and [14] first segment the breast into fatty and non-fatty regions followed by  
 139 feature extraction on each region and feature selection to remove uninforma-  
 140 tive descriptors. Several machine learning algorithms were employed by [11]  
 141 and reported an accuracy of 77%, 72% and 86% for  $k$ -Nearest Neighbours,  
 142 C4.5 and Bayesian Network classifiers, respectively. [14] reported an accu-  
 143 racy over 91% using rough fuzzy approaches with the same feature extraction  
 144 techniques used by [11]. [13] selected a set of first- and second-order statis-  
 145 tical features using a combination of several feature selection techniques and  
 146 reported a maximum accuracy of 79.2%. [12] extracted features from the

147 Spatial Grey Level Dependency matrix (a variant of GLCM) as dense tis-  
148 sue descriptors and employed a combined classifier paradigm based on feed-  
149 forward Artificial Neural Network (ANN) models to achieve just over 71%  
150 classification accuracy.

151 Morphology-based descriptors also have been widely investigated to char-  
152 acterise the shape appearance of fatty and non-fatty regions. In many cases,  
153 they are used as a set of descriptors with first- and second-order statistical  
154 features. For example, [11] and [14] extracted a set of morphology features  
155 such as the area, mean, standard deviation and kurtosis using a moment  
156 based histogram technique from the segmented fatty and non-fatty regions.  
157 The fractal-based feature extraction technique based on a set of threshold  
158 values is often used to measure the ratio between the number of low and  
159 high grey-levels within a region of interest (e.g., within a  $5 \times 5$  neighbor-  
160 hood). For example, [15] divided breast regions by recursively splitting it  
161 into quadrants depending on the intensity information that is contained in  
162 each region. Subsequently, each subregion is evaluated until the decision  
163 function determines that the region does not need to be split, thus obtain-  
164 ing regions with similar properties of tissue. They reported an accuracy of  
165 93% based on two-class classification (low versus high risk). [16] used fractal  
166 analysis to characterise parenchymal patterns in digital mammograms for  
167 risk assessment of developing breast cancer, and reported an area under the  
168 curve value of 0.86.

169 Bag-Of-Words features such as Textons and Scale Invariant Feature Trans-  
170 form (SIFT) have also been studied in the last decade. [17] compared the  
171 performance of Textons (extracted from the original images instead of the fil-  
172 tered images) and SIFT features for four-class density classification and used  
173 probabilistic Latent Semantic Analysis (pLSA) to obtain a meaningful set  
174 of features generating a compact tissue representation of each density class.  
175 They reported over 91% and 88% accuracies achieved by Textons and SIFT  
176 approaches, respectively. In a similar study, [18] compared the performance  
177 of various methods for mammographic dense tissue pattern modelling such  
178 as the Local Grey-Level appearance (LGA), Textons, LBP and Basic Image  
179 Features (BIF). In the Textons approach, [18] used various filter sizes and  
180 orientations to extract texture information from filtered images instead of  
181 directly from the original images and various parameter values (e.g., window  
182 size and the number of neighbours) were tested to extract LGA, LBP, and  
183 BIF features. For the four-class BI-RADS classification the authors reported  
184 accuracies 75%, 72%, 59% and 70% for Textons, LGA, LBP and BIF, re-

185 spectively. [19] implemented a Textons approach based on the Maximum  
186 Response 8 (MR8) filter bank. The  $\chi^2$  distribution was used to compare  
187 each of the resulting histograms from the testing set with all the learned  
188 histogram models from the training set and reported 75.5% accuracy.

189 In recent studies, [20, 21] showed that both LTP (three-encoding ap-  
190 proach) and LQP (five-encoding approach) operators can achieve 77% to 85%  
191 accuracy using a multiscale approach with different neighbourhood topolo-  
192 gies and different number of dominant patterns. Instead of extracting texture  
193 features from the entire breast region, [20, 21] showed that extracting texture  
194 information only from the Corpus Mammea reduces the number of overlap-  
195 ping features and hence produces more distinctive features. The authors  
196 showed that this approach was at least 7% better in terms of classification  
197 accuracy compared to classifying breast density by extracting features from  
198 the entire breast region. [22] modelled the distribution of breast density based  
199 on the multi-scale distribution of dark ellipses (representing fatty tissue) and  
200 bright ellipses representing fibroglandular and dense tissues. Subsequently,  
201 morphological features such as distances, sizes and diameters were extracted  
202 to estimate the amount of dense tissue in the breast. Preliminary results  
203 show that their method achieved accuracy between 67% to 72% based on  
204 various classifiers. Later, [23] compared the performance between Elliptical  
205 LBP (ELBP), uniform ELBP and Mean-Elliptical LBP (M-ELBP) for three-  
206 class classification (fatty, glandular and dense) and reported classification  
207 accuracies of 75%, 74% and 80%, respectively.

208 Despite a large number of methods having been developed to address  
209 the breast density classification problem in mammograms, only a few studies  
210 have achieved accuracies above 80% and the majority of the methods pro-  
211 duced between 70% to 79%. Moreover, most studies were based on two-class  
212 classification (low risk (BI-RADS I and II) versus high risk (BI-RADS III and  
213 IV)) or three-class classification (fatty, fatty-glandular and dense-glandular)  
214 instead of four-class classification. This might be due to challenging issues  
215 such as complex and overlapping tissue appearance as well as ambiguous tex-  
216 ture patterns which make it difficult to separate BI-RADS classes based on  
217 texture.

218 In the last a few years, Convolutional Neural Network (CNN) and Deep  
219 Residual Learning (ResNet) are among the most popular deep learning tech-  
220 niques which have been used to address classification and segmentation prob-  
221 lems in clinical image data [8, 7]. Many current studies have claimed that  
222 deep learning based methods produced superior results achieving accuracy

223 similar to human performances [8, 7, 24, 25]. In breast density classifica-  
 224 tion, the majority of the studies used deep learning based methods to address  
 225 either two-class (scattered and heterogeneously density) or three-class clas-  
 226 sification problems (fatty, glandular and dense tissue). [26] developed a  
 227 CNN model to distinguish between scattered density and heterogeneously  
 228 dense tissues using over 22,000 images. [27] used deep learning networks  
 229 to classify dense and non-dense samples for the purpose of dense tissue seg-  
 230 mentation. [28] used unsupervised deep learning to segment dense tissue  
 231 in mammograms and estimate the risk of developing breast cancer based on  
 232 the of segmentation result. Recently, [29] reported that their deep learning  
 233 classification results are correlated well with BI-RADS density assessments  
 234 by radiologists and comparable with a state of the art algorithm, Laboratory  
 235 for Individualized Breast Radiodensity Assessment (LIBRA).

236 Although deep learning based methods have shown promising results,  
 237 breast density classification based on the BI-RADS four-class assessment  
 238 guideline (fourth edition) remains a challenging task. This might be due to  
 239 a lack of annotated data and the complexity of the task itself. According  
 240 to the recent study of [8], the main challenge of employing a deep learning  
 241 network is the requirement of a large dataset and annotations from experts.  
 242 In addition, deep learning suffers from a lack of direct human interpretabil-  
 243 ity because deep learning features rely on filter responses solicited from a  
 244 large amount of training data whereas hand-crafted features such as those  
 245 extracted from local patterns provide transparent information, which are  
 246 more interpretable to clinicians and researchers. Furthermore, all datasets  
 247 available publicly suffer from an imbalance in the number of images for each  
 248 class. For example, in the MIAS [10] and InBreast [30] datasets the number  
 249 of BI-RADS IV cases is less than 10% of the total number of cases. As a  
 250 result, deep learning networks may suffer from an imbalanced classification  
 251 problem.

## 252 5. Methodology

253 This section explains the technical details of our study covering steps  
 254 involved in the pre-processing, feature extraction and the classification phase.  
 255 Figure 3 shows a general overview of the workflow in our experiments.

256 Firstly, we performed breast segmentation to separate the breast bound-  
 257 ary and pectoral muscle using the method proposed by [31]. This ensures  
 258 that only local patterns within the breast region will be extracted. Follow-

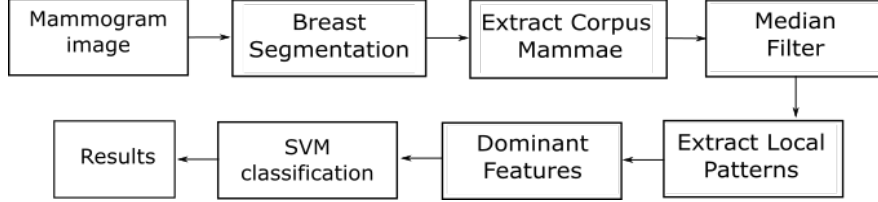


Figure 3: An overview of the work flow involved in our experiments.

259 ing our previous studies [20, 21], we extract the Corpus Mammae region of  
 260 the breast to reduce the risk of overlapping local patterns across BI-RADS  
 261 classes. Subsequently, we used a median filter around  $3 \times 3$  neighbouring pix-  
 262 els for noise reduction. We extract local patterns at different channels using  
 263 the LBP, LTP, LQP and LSP operators within the estimated Corpus Mam-  
 264 mae region. We select a set of dominant patterns to remove uninformative  
 265 features and employ the Support Vector Machine (SVM) as a classification  
 266 approach after a zero-mean normalisation.

### 267 5.1. Pre-processing

268 To segment the breast region, we used our previously developed method  
 269 [31] to estimate the breast and pectoral muscle boundaries. Based on the  
 270 estimated boundaries we created a breast mask and performed a pixel wise  
 271 multiplication with the original image resulting in an image with only the  
 272 breast region. The method [31] employs a region-based Active Contour  
 273 to estimate the breast boundary and restricted contour growing with edge  
 274 information for the pectoral muscle boundary estimation. Figure 4 shows an  
 275 example of an extracted Corpus Mammae region of interest ( $ROI_{cm}$ ) and a  
 276 few examples of estimated breast and pectoral muscle boundaries.

277 The left-most image in Figure 4 shows the estimated  $ROI_{cm}$  area (amber  
 278 square box). To extract  $ROI_{cm}$ , we find the height( $B_h$ ) and the width of the  
 279 breast ( $B_w$ ).  $B_h$  is then relocated to the middle of  $B_w$  to get the intersection  
 280 point. The width and height of the square area of the  $ROI_{cm}$  (amber line  
 281 Figure 4) can be computed as  $B_w \times B_w$  with the centre located at the inter-  
 282 section point between the  $B_h$  and  $B_w$  lines.  $B_h$  is the height of the breast,  
 283 which is the longest perpendicular distance between the  $x$ -axis and the breast  
 284 boundary. Note that the size of  $ROI_{cm}$  varies depending on the width of the  
 285 breast and local patterns are only extracted within  $ROI_{cm}$  instead of from  
 286 the whole breast region as the majority of the studies in the literature have

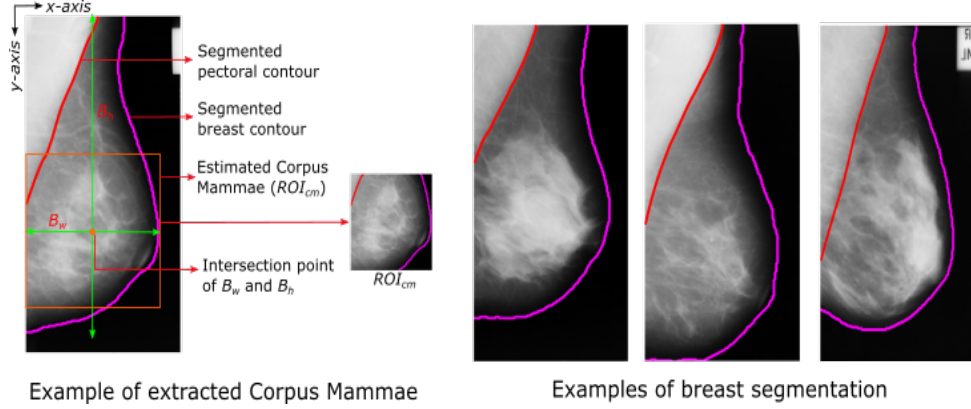


Figure 4: Left: an example of extracted Corpus Mammae. Right: examples of breast segmentations.

done.

## 5.2. Feature Extraction

This section explains the original LBP operator [32] and its extensions which are the LTP and LQP operators introduced in [33] and [34], respectively. Subsequently, we propose a LSP operator and explain the different channels in each of the operators. Note that there are several other variants of the LBP operator such as ELBP [35], M-ELBP [36], Median Robust Extended Local Binary Pattern (MRELBP) [37], spatio-temporal LBP (STLBP) [38] and completed LBP (CLBP) [39]. However, they are outside the scope of our study because these operators are based on the LBP's two-channel encoding technique (nevertheless we consider this as part of our future work) whereas the scope of our study is to investigate the effects of local patterns on accuracy when various different channel encoding techniques are employed.

All these operators use three different mapping tables namely uniform pattern ('u2'), rotation invariant pattern ('ri') and both uniform and rotation invariant patterns ('riu2') to extract uniform, rotation invariant and rotation invariant uniform patterns, respectively. In our study, we investigate the 'riu2' patterns (a combination of 'u2' and 'ri' patterns) as they provide more discriminant features based on our previous studies [20, 21] and the study of



307 [32]. The value of the LBP code of the pixel  $(i, j)$  is given by:

$$LBP_{(P,R)}(i, j) = \sum_{p=0}^{(P-1)} s(g_p - g_c) 2^p \quad (1)$$

308 where  $R$  and  $P$  are the radius of the circle that forms the neighbourhood of  
 309 the operator and the number of pixels in the neighbourhood, respectively.  
 310 The grey level value of the centre pixel is denoted  $g_c$ , and  $g_p$  is the grey level  
 311 value of the  $p^{th}$  neighbour. The LBP operator thresholds the neighbouring  
 312 pixels using a two-value encoding system as shown in Equation 2.

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

313 Later, [33] introduced a three-value encoding technique (LTP operator)  
 314 which thresholds the neighbouring pixels based on a constant threshold set by  
 315 the user ( $\tau_1$ ) as shown in Equation 3. Once the LTP code image is generated,  
 316 it can be separated into two binary patterns from its positive and negative  
 317 channels. Therefore, it encodes an image into three channels but producing  
 318 two binary patterns.

$$s(x) = \begin{cases} 1, & x > g_c + \tau_1 \\ 0, & g_c - \tau_1 < x < g_c + \tau_1 \\ -1, & x < g_c - \tau_1 \end{cases} \quad (3)$$

319 In [34] the LQP operator was introduced, which thresholds the neigh-  
 320 bouring pixels using a five-value encoding technique (see Equation 4) based  
 321 on two constant thresholds  $\tau_1$  and  $\tau_2$ . Subsequently, the LQP code image is  
 322 split into four binary patterns by considering its positives, zero and negative  
 323 components. Therefore, the LQP operator encodes an image into five chan-  
 324 nels but results four binary patterns (from two positive channels ( $s(1)$  and  
 325  $s(2)$ ), zero channel ( $s(0)$ ) and combined negative channels ( $s(-1) \cap s(-2)$ )).

$$s(x) = \begin{cases} 2, & x \geq g_c + \tau_2 \\ 1, & g_c + \tau_1 \leq x < g_c + \tau_2 \\ 0, & g_c - \tau_1 \leq x < g_c + \tau_1 \\ -1, & g_c - \tau_2 \leq x < g_c - \tau_1 \\ -2, & \text{otherwise} \end{cases} \quad (4)$$

Figure 5 shows the resulting code images and local patterns from each channel using the LBP, LTP and LQP operators. A histogram from the LBP code image ( $H_{lbp}$ ) can be extracted to characterise local patterns in  $ROI_{cm}$ . However, for LTP and LQP code images, local patterns can be extracted from each channel. For example, local patterns that represent  $ROI_{cm}$  when using the LTP operator can be extracted by computing the histogram image of ‘LTP Channel 1’ and ‘LTP Channel 2’, donated as  $H_{ltp1}$  and  $H_{ltp2}$ , respectively and concatenating these histograms as the final feature descriptors. Similarly, to extract local patterns in  $ROI_{cm}$  using LQP operator, we can compute and concatenate the histograms from all LQP channels ( $H_{lqp1}$ ,  $H_{lqp2}$ ,  $H_{lqp3}$  and  $H_{lqp4}$ ). Note that all channel images are binary images. As a result of histogram concatenating more local patterns can be extracted using the LTP and LQP operators.

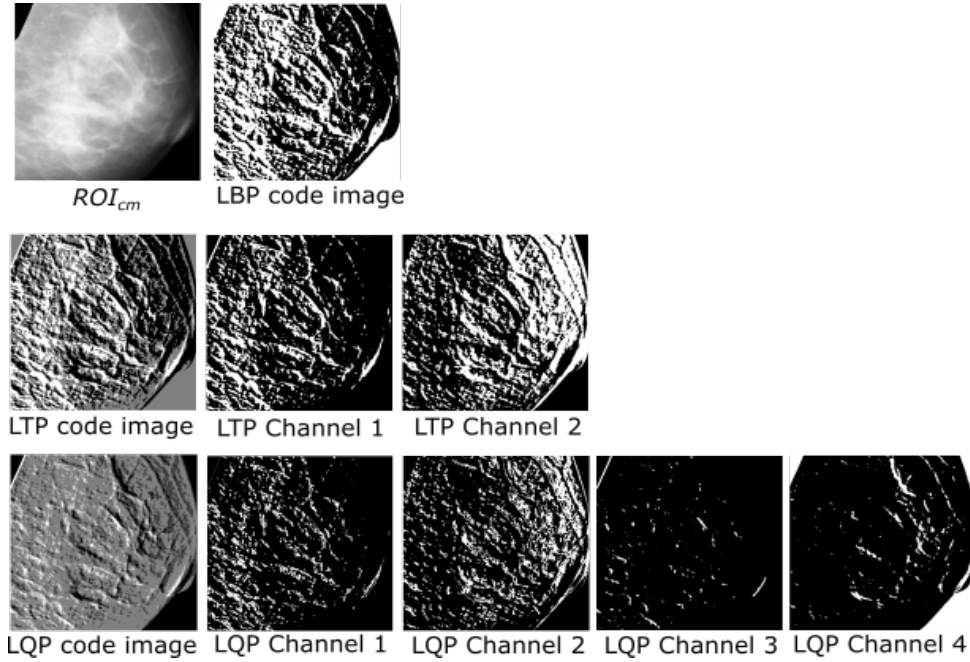


Figure 5: Examples of LBP, LTP and LQP code images and local pattern images from different channels.

We took a similar approach by thresholding the neighbouring pixels using a seven-value encoding technique (hence the name is septenary) based on three threshold values ( $\tau_1$ ,  $\tau_2$  and  $\tau_3$ ). The value of the LSP code of the pixel

342  $(i, j)$  is given by:

$$LSP_{(P,R)}^{pattern}(i, j) = \sum_{p=0}^{(P-1)} s_{pattern}(g_p) 2^p \quad (5)$$

343 where  $pattern \in \{1, 2, 3, 4, 5, 6\}$  represents six binary patterns by considering  
 344 its upper-positive, middle-positive, lower-positive, upper-negative, middle-  
 345 negative and lower-negative components denoted as  $s(3)$ ,  $s(2)$ ,  $s(1)$ ,  $s(-1)$ ,  
 346  $s(-2)$  and  $s(-3)$ , respectively, as shown in Equation 6

$$s(x) = \begin{cases} 3, & x \geq g_c + \tau_3 \\ 2, & g_c + \tau_2 \leq x < g_c + \tau_3 \\ 1, & g_c + \tau_1 \leq x < g_c + \tau_2 \\ 0, & g_c - \tau_1 \leq x < g_c + \tau_1 \\ -1, & g_c - \tau_2 \leq x < g_c - \tau_1 \\ -2, & g_c - \tau_3 \leq x < g_c - \tau_2 \\ -3, & \text{otherwise} \end{cases} \quad (6)$$

347 Note that LSP encodes an image into seven channels but results in six binary  
 348 patterns. The LSP code image is split into six binary patterns using the  
 349 following conditions

$$s_1(x) = \begin{cases} 1, & \text{if } s(3) = 3 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$s_2(x) = \begin{cases} 1, & \text{if } s(x) = 2 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

$$s_3(x) = \begin{cases} 1, & \text{if } s(x) = 1 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

$$s_4(x) = \begin{cases} 1, & \text{if } s(x) = -1 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

$$s_5(x) = \begin{cases} 1, & \text{if } s(x) = -2 \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

$$s_6(x) = \begin{cases} 1, & \text{if } s(x) = -3 \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

Figure 6 shows an example of a LSP code image extracted from  $ROI_{cm}$  in Figure 5 and its local patterns from different channels. Using the encoding technique in Equation 6, we can generate the LSP code image of  $ROI_{cm}$ . The LSP code image is split into six binary images which represent local patterns in six channels. Unlike LTP and LQP where the user has to manually determine threshold values, we introduce an automatic approach by computing the number of neighbours with grey level value  $\leq 25^{th}$  percentile of the entire neighbourhood, number of neighbours with grey level value between  $25^{th}$  and  $75^{th}$  percentile of the entire neighbourhood and number of neighbours with grey level value  $\geq 75^{th}$  percentile. Subsequently, we sort the values in ascending order for  $\tau_1$ ,  $\tau_2$  and  $\tau_3$  values.

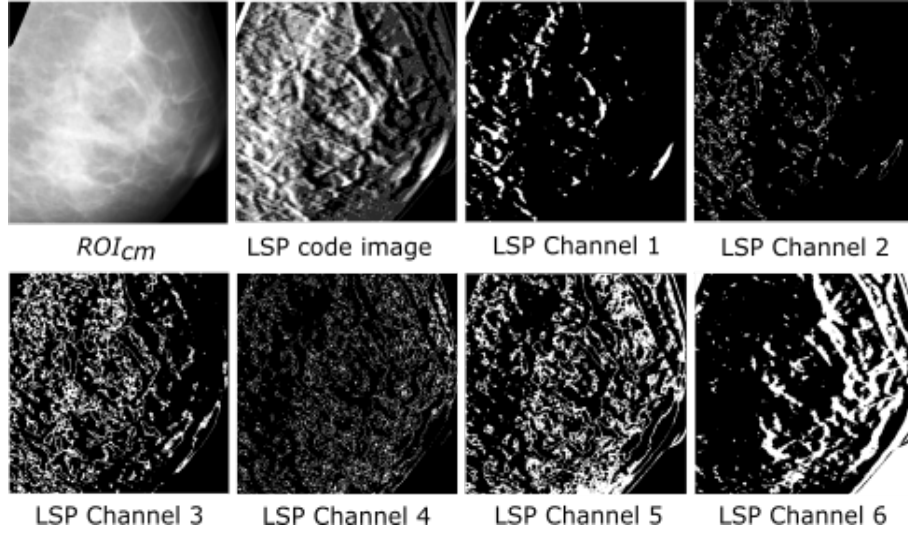


Figure 6: Examples of LSP code image and its local pattern images from different channels.

The LBP, LTP, LQP and LSP are similar in terms of architecture as each is defined using a circle centred on each pixel and the number of neighbours. The main difference is the LSP thresholds the neighbouring pixels into seven (3, 2, 1, 0, -1, -2 and -3) values. In contrast, the LBP, LTP and LQP encode an image into two (1 and 0), three (1, 0 and -1), and five (2, 1, 0, -1 and -2) values. Figure 7 shows an example of the feature extraction process. The final

367 histogram is a concatenation of six histograms computed from each binary  
 368 pattern (binary image, e.g., LSP code 1) generated based on the conditions  
 369 in equations (7) to (12). The process of capturing local patterns of  $ROI_{cm}$  is  
 370 also the same for the other operators such as the LBP, LTP and LQP, except  
 371 less number of channels (resulting in a shorter concatenated histogram).

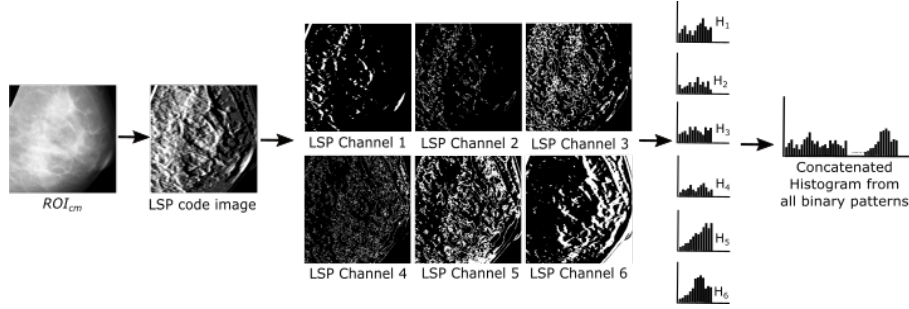


Figure 7: Summary of feature extraction phase using the LSP operator. Note that the binary pattern from each channel is computed from the LSP code image and a histogram image is computed for each channel. Finally, all histograms are concatenated and treated as a feature vector to represent the local pattern of  $ROI_{cm}$ .

### 372 5.3. Dominant Patterns

373 Following our previous studies in [21, 20], we selected dominant patterns  
 374 to reduce feature dimension, hence simplifying the prediction model. Domi-  
 375 nant patterns are patterns that occur frequently in the training set [40]. Let  
 376  $I_1, I_2 \dots I_j$  be images in the training set. We compute the histogram feature  
 377 ( $H_{I_j}^{LSP}$ ) for each training image and perform a bin-wise summation for all  
 378 the histograms to find the pattern's distribution from the training set. Sub-  
 379 sequently, the resulting histogram ( $H^{LSP}$ ) is sorted in descending order, and  
 380 the patterns corresponding to the first  $D$  bins are selected using the following  
 381 equation:

$$D = \arg \min_N \frac{\sum_{i=1}^{N-1} H^{LQP}(i)}{\sum_{i=1}^{2^P} H^{LQP}(i)} > 0.01 \times n \quad (13)$$

382 where  $N$  and  $n$  are the total number of patterns and the threshold chosen  
 383 by the user, respectively. For example choosing  $n = 98$  means removing  
 384 patterns that occur with relative frequency of less than 2% in  $H^{LSP}$ . The  
 385 smaller the value of  $n$ , the smaller the number of patterns selected.

#### 386 5.4. Classification

387 Once the feature extraction is completed, we employed several machine  
 388 learning algorithms in WEKA [41] to build our classification models. For  
 389 machine learning with only one parameter (e.g.,  $k$ -NN), the CVPParameterS-  
 390 election technique was employed. In contrast, the GridSearch technique was  
 391 used to explore two parameters for classifiers with two parameters (e.g. RF  
 392 and SVM). The CVPParameterSelection finds the best setting based on op-  
 393 tions set by the user by optimising the classification accuracy. The Grid-  
 394 Search tests all possible combination of two parameters and selects the best  
 395 combinations based on the highest accuracy. To find the best parameters for  
 396 each classifier, 25 patients (9, 7, 5, 4 patients for BI-RADS I, II, III and IV,  
 397 respectively) were selected from the MIAS dataset and 3-fold cross-validation  
 398 was used to evaluate the performance for each (or pair) of the tested param-  
 399 eter(s) during the optimisation process. The classifier was trained, and in  
 400 the testing phase, each unseen  $ROI_{cm}$  from the testing set is classified as  
 401 BI-RADS I, II, III or IV. The classifiers used in this study are presented in  
 402 Table 1.

Table 1: List of classifiers used in this study. The # sign represents number

Classifiers	Default parameters in WEKA
Support Vector Machine (SVM)	Complexity ( $C$ )=1.0, exponent=1
Random Forest (RF)	# of forests=100, depth = 0 (unlimited)
Multilayer Perceptron (MLP)	Learning rate=0.3, momentum=0.2
$k$ -Nearest Neighbours ( $k$ -NN)	$k$ =1, Euclidean distance

#### 403 6. Experimental Results

404 To test the performance of the method we used the following datasets:  
 405 (a) the Mammographic Image Analysis Society (MIAS) dataset [10] which  
 406 consists of 322 mediolateral-oblique (MLO) mammograms of 161 women.  
 407 Films were taken from the UK National Breast Screening Programme and  
 408 have been digitised to  $50\mu\text{m} \times 50\mu\text{m}$  and quantised to 8 bits. The distribution  
 409 for BI-RADS classes is 60 (BI-RADS I), 105 (BI-RADS II), 129 (BI-RADS  
 410 III) and 31 (BI-RADS IV), and (b) the InBreast dataset [30] which consists  
 411 of 206 MLO mammograms from 103 patients. Each image is direct digital  
 412 acquisition on a-Se imaging plates. The pixel size of all images is  $70\mu\text{m} \times$

70  $\mu\text{m}$ , with 14-bit contrast resolution. The density distribution for the BI-RADS classes is as follow: 69 (BI-RADS I), 74 (BI-RADS II), 49 (BI-RADS III), and 14 (BI-RADS IV). Each image contains BI-RADS information (e.g., BI-RADS class I, II, III or IV) provided by an expert radiologist based on the fourth edition of the BI-RADS system. We ran stratified 5-fold cross validation for 10 times.

Accuracy ( $Acc$ ) is used to measure the performance of the method, which represents the total number of correctly classified images as a proportion of the total number of images.

### 6.1. Optimised Parameters

Table 2 shows a list of parameter values tested and optimised values for the four classifiers employed in this study. Note that, the parameter optimisation was performed based on 25 patients (each BI-RADS class has six or seven patients) taken from the MIAS dataset. Subsequently, we use these parameter values in the testing phase for both MIAS and InBreast datasets. However, for the MIAS dataset we only use 136 patients (272 mammograms) and we excluded 25 patients (50 mammograms) which were used for parameter optimisation. From now on, all parameter settings for the classifiers employed in this study are based on the best parameter values summarised in Table 2.

Table 2: List of parameter options tested in this study.

Classifiers	Parameter tested	Best parameters
SVM	Kernel = 'Polynomial', $C = 1$ to 10 $e = 1$ to 5	$C = 5$ , $e = 1$
RF	Number of forests ( $rF$ )=1 to 165 depth ( $D$ )=0 to 10	$rF = 70$ , $D = 0$
MLP	Learning rate ( $LR$ )=0.1 to 2.0 (e.g. 0.1, 0.2 ... 2.0) momentum( $M$ )=0.1 to 2.0 (e.g. 0.1, 0.2 ... 2.0)	$LR = 0.1$ , $M = 0.5$
$k$ -NN	$k$ =1-49 (e.g. 1, 3, 5, ... 49) Euclidean distance	$k = 5$

### 6.2. Quantitative Results

This section presents classification results for LBP, LTP, LQP and LSP operators based on different classifiers. Since the LTP and LQP require

threshold values from a user, the parameters  $\tau_1$  and  $\tau_2$  were selected based on previous studies [21, 20]. Therefore the  $\tau_1$  value for the LTP operator is set to 5 and  $\tau_1$  and  $\tau_2$  values for LQP are set to 5 and 12, respectively. Note that the average *Acc* represents the mean accuracy across different numbers of dominant patterns ( $n$ ) where the maximum *Acc* is the highest accuracy achieved for  $n$  number of dominant patterns. Choosing  $n = 93$  means removing patterns that occur with a frequency of less than 7% in the histogram feature. Numerical values represent performance evaluation based on 272 and 206 images for the MIAS and InBreast datasets, respectively. Note that we have used 50 images (of the original 322 images) from the MIAS dataset for parameter optimisation.

Figure 8 and 9 shows classification results using LBP operator for the MIAS and InBreast dataset, respectively. It can be observed that for the MIAS dataset a maximum accuracy of 73.8% is achieved by  $LBP(7, 16)$  (where  $R = 7$ ,  $P = 16$ , e.g.  $LBP(R, P)$ ) using the SVM classifier at  $n = 91.5$  (removing local patterns with frequency of less than 8.5% in the histogram features). The best average accuracy of 69.7% is achieved when employing  $LBP(7, 16)$  with  $n = 92.1$ . For the InBreast dataset a maximum accuracy of 73.8% is achieved by  $LBP(5, 10)$  at  $n = 96.5$  whereas best average accuracy (70.2%) is achieved by  $LBP(9, 20)$  using the MLP classifier. Overall, the majority of the classification accuracies fall in a range between 65% to 70%.

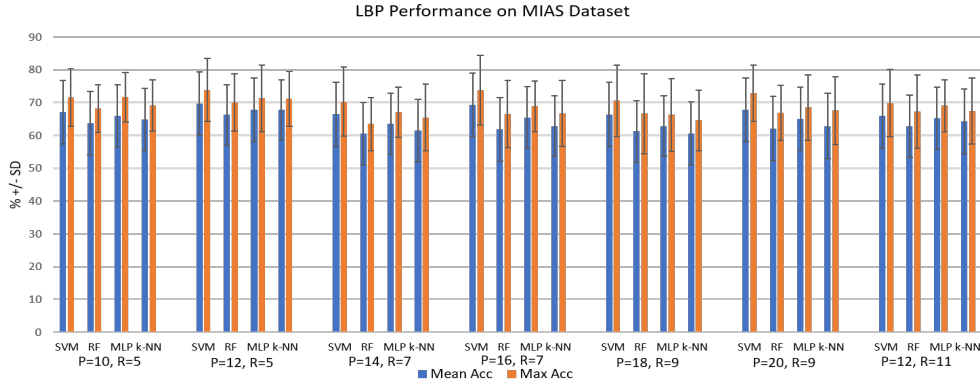


Figure 8: LBP performance on MIAS dataset

Figure 10 and 11 shows quantitative results using the LTP operator for the MIAS and InBreast dataset, respectively. The majority of the average accuracies are in a range between 70% to 75% which are higher than the



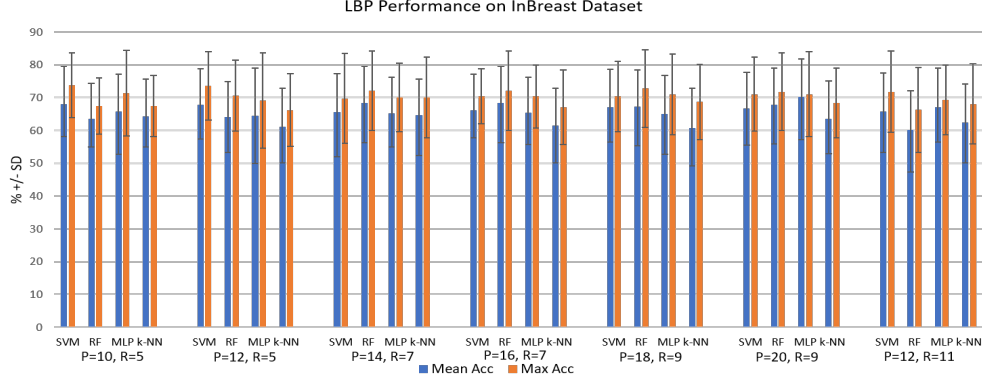


Figure 9: LBP performance on InBreast dataset

ones produced by the LBP operator; for maximum accuracy, many cases are over 75%. Furthermore, it can be observed that the LTP operator outperformed the LBP operator regardless of the  $R$  and  $P$  values. The highest accuracies are 81% (average 77.5%) and 78.7% (average 74.3%) on the MIAS and InBreast datasets, respectively. They are at least 4% to 7% higher than the maximum accuracy produced when extracting local patterns using the LBP operator. The SVM classifier produced the highest classification accuracy by removing 6.5% of the local patterns when evaluated on the MIAS dataset ( $LTP(7, 14)$  and  $LTP(7, 16)$ ) where the MLP classifier ( $LTP(9, 18)$ ) outperformed the other classifiers at  $n = 93.1\%$  when tested on the InBreast dataset.

Figure 12 and 13 shows classification results using the LQP operator for the MIAS and InBreast dataset, respectively. Overall, it can be observed that LQP produced better classification results in comparison to LBP and LTP operators. The majority of the maximum accuracies are in a range between 74% to 79% which indicates that local patterns extracted by the LQP operator are more discriminant in comparison to the previous two operators.  $LQP(9, 18)$  and  $LQP(9, 20)$  produced a maximum accuracy of 82.1% with the best average accuracy of 78.6% at  $n = 90$  on the MIAS dataset. However, when evaluated on the InBreast dataset the maximum accuracy of 80.1% is achieved by  $LQP(9, 18)$  at  $n = 98$  and the best average accuracy is achieved by  $LQP(9, 20)$ . Moreover, it can also be observed that the SVM classifier produced an average maximum accuracy over 80% regardless of the values of  $P$  and  $R$  when evaluated on the MIAS dataset. However, the MLP classifier

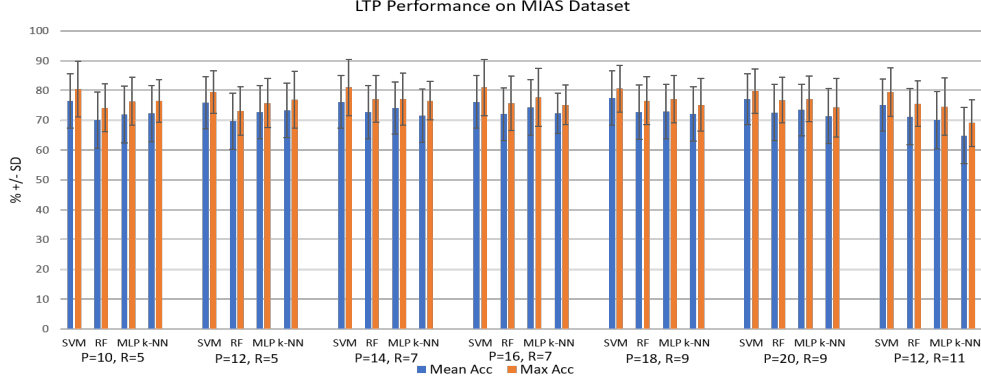


Figure 10: LTP performance on MIAS dataset

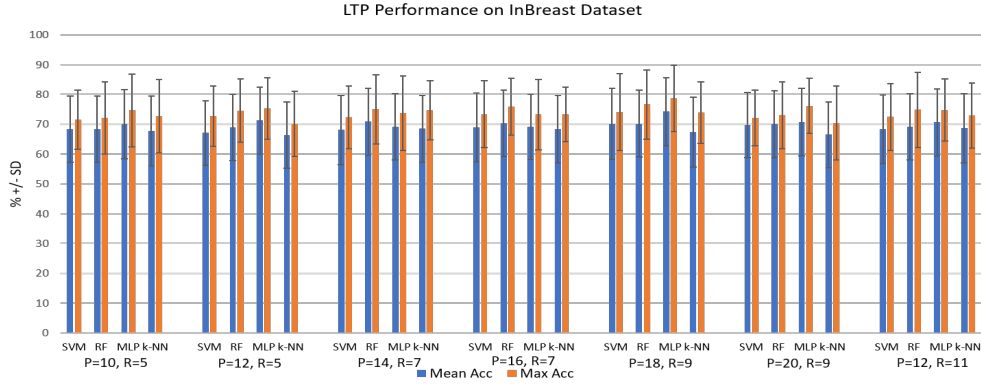


Figure 11: LTP performance on InBreast dataset

484 tends to produce better results when tested on the InBreast dataset with the  
 485 best average  $Acc = 75.5\%$  which is 1% better than the SVM classifier.

486 Table 14 and 15 shows quantitative results using our proposed LSP op-  
 487 erator when evaluated on the MIAS and InBreast dataset, respectively. The  
 488 majority of the classification accuracies fall in the range 75% to 80% which  
 489 is slightly better compared to the results produced by the LQP operator.  
 490 Experimental results show that the LSP operator produced a maximum accu-  
 491 racy of 83.3% (which outperforms the LBP (73.8%), LTP (81%) and LQP  
 492 (82.1%) operators) on the MIAS dataset. The best average accuracy of 81.6%  
 493 is achieved by  $LSP(9, 18)$  at  $n = 99.1$  which is at least 2.8% higher than the

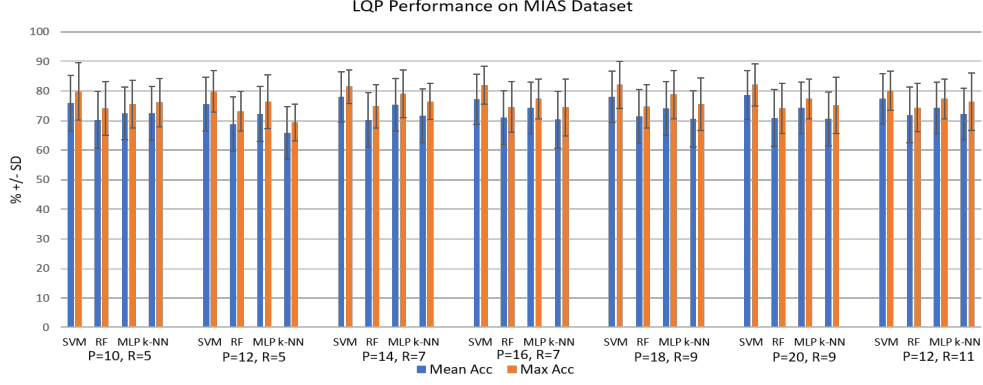


Figure 12: LQP performance on MIAS dataset

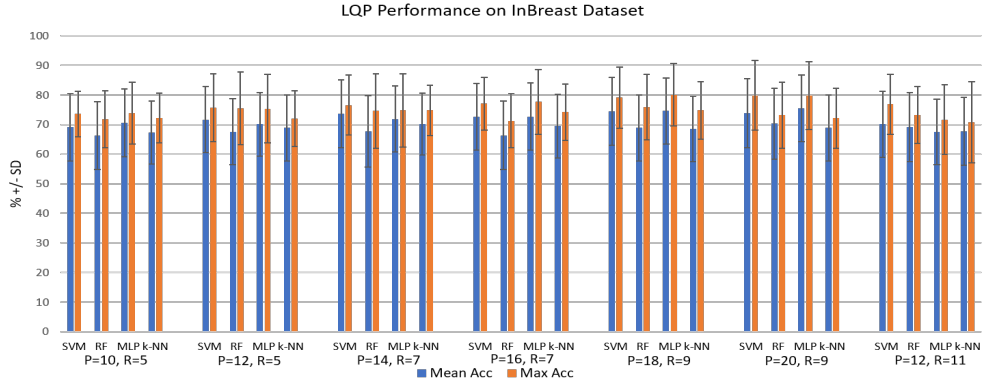


Figure 13: LQP performance on InBreast dataset

best average  $Acc$  produced by LQP and LTP. The LSP operator achieved maximum 80.5% classification accuracy at  $n = 95.5$  using the SVM classifier whereas the highest accuracy achieved by RF,  $k$ -NN and MLP are 77.9 ( $LSP(11, 22)$ ), 79.7% ( $LSP(9, 20)$ ) and 79.8 ( $LSP(7, 16)$ ), respectively. Regarding the best average accuracy across different numbers of dominant patterns  $LSP(9, 20)$  produced 77.1% using the SVM classifier. RF,  $k$ -NN and MLP produced 70.9%, 73.3% and 73.7%, respectively when  $LSP(9, 20)$  is employed. Overall, the LSP operator produced more discriminant local patterns in separating the BI-RADS classes in both MIAS and InBreast datasets. We will discuss this further in the subsequent subsections.

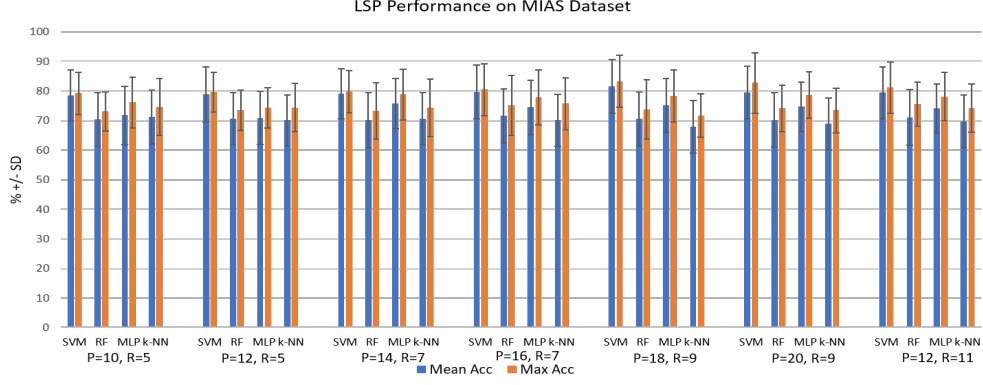


Figure 14: LSP performance on MIAS dataset

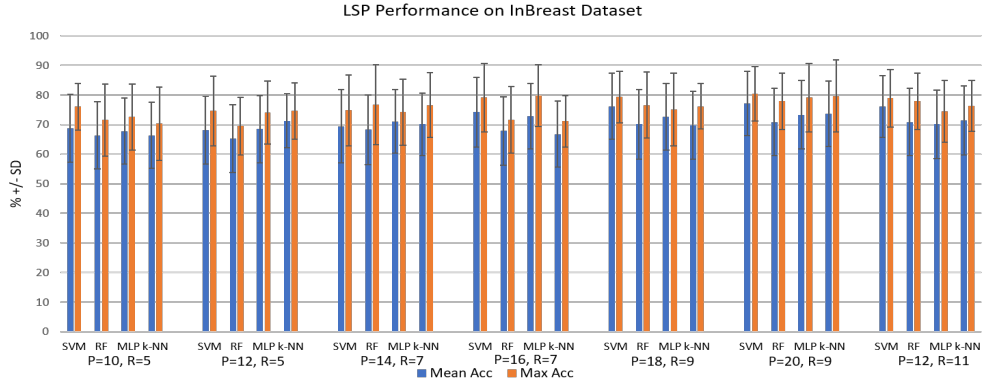


Figure 15: LSP performance on InBreast dataset

### 6.3. Maximum and Average Performance Comparison Between Operators

This section summarises the maximum and average results produced by the operators across different classifiers based on the MIAS and InBreast datasets. For the MIAS dataset, Figures 16 and 17 show performance comparisons between LBP, LTP, LQP and LSP operators using different classification approaches. In terms of maximum accuracy, it can be observed that LSP ( $Acc = 83.3\%$ ) outperformed the other texture operators (LBP, LTP and LQP) when using SVM, RF and MLP as classification approaches. However, when the  $k$ -NN classifier is employed, the maximum classification accuracy is  $76.9\%$  using LTP which is  $0.7\%$  and  $1.2\%$  better than the LQP

and LSP operators, respectively. Regarding the average accuracies across different numbers of dominant patterns, both the SVM and MLP classifiers produced higher results when using LQP and LSP features. For example, the LSP operator produced 81.6% and 75.8% accuracies for SVM and MLP classifiers, respectively compared to 78.6% (SVM) and 75.4% (MLP) when classification was performed based on local patterns extracted using the LQP operator. In contrast, classification accuracy is at least 0.9% better for local patterns using LTP when RF and  $k$ -NN classifiers are used compared to LQP and LSP. In both Figures 16 and 17, the LBP operator produced the lowest accuracy results regardless to classification approach which clearly shows that a channel encoding technique is necessary to capture more discriminant features and hence improve the classification results.

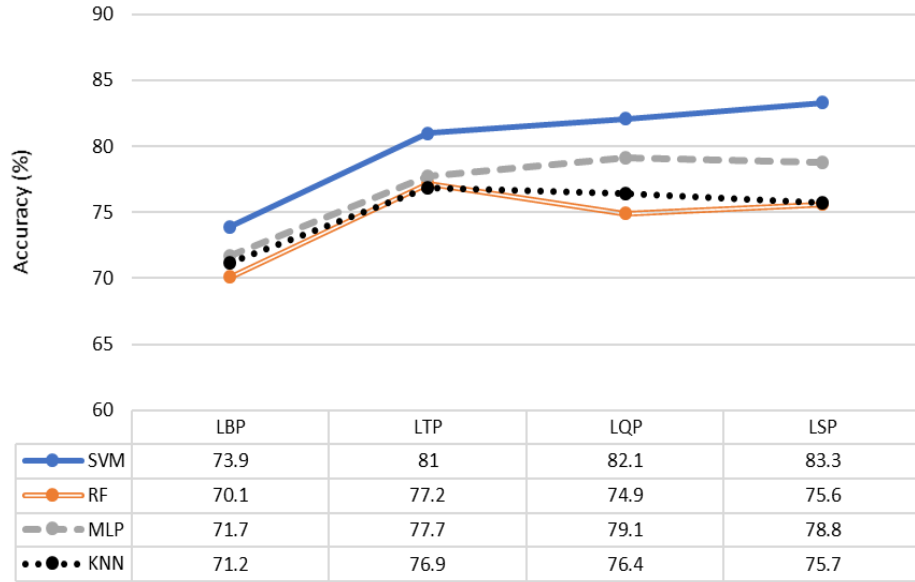


Figure 16: Comparison of maximum accuracies achieved by each operator across different classifiers when evaluated on the MIAS dataset.

For the InBreast dataset, Figures 18 and 19 show performance comparisons between all operators using different classifiers. Regarding maximum accuracy, the LSP operator achieved  $Acc = 80.5\%$  using the SVM classifier and outperformed the LBP, LTP and LQP operators by 6.7%, 6.4%, and 0.6%, respectively. An improvement also can be observed for the  $k$ -NN classifier as the LSP operator produced 79.7% compared to the other operators

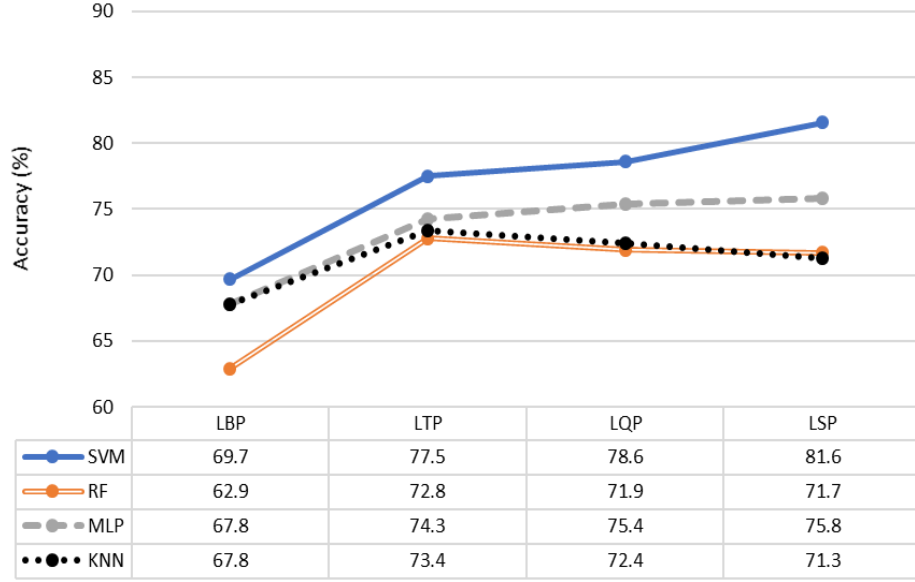


Figure 17: Comparison of average accuracies across different numbers of dominant patterns achieved by each operator across different classifiers when evaluated on the MIAS dataset.

532 which produced  $Acc < 75\%$ . For the RF classifier, the LSP operator pro-  
 533 duced 77.9%, at least 1.2% higher compared to the other operators. The  
 534 LQP operator produced a maximum accuracy of 80.1% which is 0.3% higher  
 535 than the LSP operator when using the MLP classifier. Once again it can  
 536 be observed that other operators always outperform the classification results  
 537 of the LBP operator regardless of classifiers. In terms of average accuracies,  
 538 the LSP operator once again outperforms the other operators when the SVM  
 539 classifier is employed with the best average accuracy of 77.1%. Similar trends  
 540 can be seen when using the  $k$ -NN classifier where the LSP features produced  
 541 the best average accuracy of 73.8% compared to 64.6%, 68.7% and 70.2%  
 542 produced by the LBP, LTP and LQP features, respectively. The LSP and  
 543 LTP operators produced similar results of 70.9% when the RF classifier is  
 544 employed which is 0.6% higher than the LQP operator. The LSP operator  
 545 produced the best average accuracy of 73.3% using the MLP classifier which  
 546 is slightly lower compared to LQP and LTP operators with 75.5% and 74.3%  
 547 accuracy, respectively. Once again experimental results suggest that the LSP  
 548 operator extracts more discriminant local features and performs better when  
 549 the SVM classifier is employed.

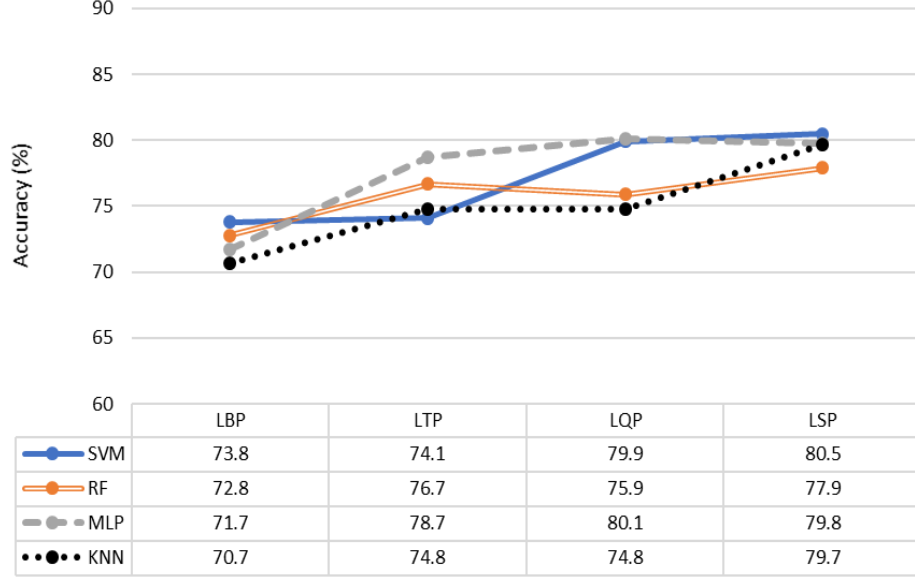


Figure 18: Comparison of maximum accuracies achieved by each operator across different classifiers when evaluated on the InBreast dataset.

## 550 7. Discussion

551 In this section we will discuss (i) the effects on accuracy when different  
552 numbers of dominant patterns are used in the classification, (ii) the effects of  
553 the radius ( $R$ ) and number of neighbours ( $P$ ) in the local patterns discrimi-  
554 natory levels; (iii) explanations on why multichannel local patterns produced  
555 more discriminant features in comparison to the original LBP operator, in  
556 (iv) extending to 9 and 11- encoding systems, (v) discussion of the existing  
557 studies in the literature, (vi) statistical analysis and (vii) future work.

### 558 7.1. Effects of Different Numbers of Dominant Patterns ( $n$ )

559 To investigate the effects of  $n$  on the classification accuracy for all oper-  
560 ators described in this study we tested 60 different values from 90 to 99.9  
561 (e.g.,  $n \in \{90, 90.1, 90.3, 90.5, \dots, 99.9\}$ ). In other words, we investigated the  
562 variation of classification accuracy by including local patterns which have  
563 a frequency of minimum 0.1% to 10% in the histogram feature (e.g. local  
564 patterns with a relative frequency of less than 0.1% to 10% are removed,  
565 resulting to a shorter histogram feature). Figure 20 shows the effects of  $n$

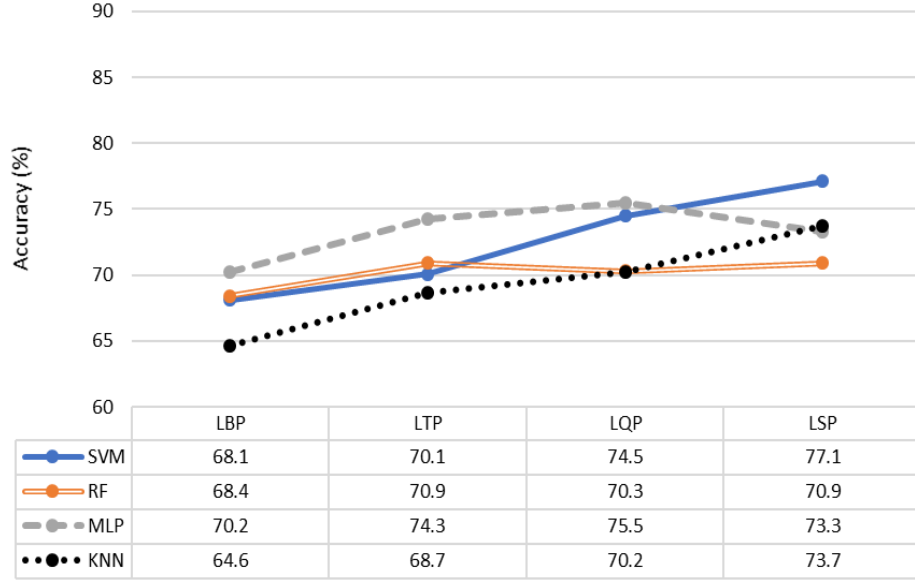


Figure 19: Comparison of average accuracies across different numbers of dominant patterns achieved by each operator across different classifiers when evaluated on the InBreast dataset.

on classification accuracy for LBP, LTP, LQP and LSP operators. In this experiment, we employed the SVM classifier as it produced better performance compared to the MLP, RF and  $k$ -NN classifiers. It can be observed that  $n$  plays an important role in getting the best classification accuracy. For the LBP operator it can be observed that a large variation of classification accuracy between 63% to 74% occurs, with a 9.71% standard deviation. The LQP operator produced a standard deviation of 9.25% which is higher than the LTP operator of 9.04%. The LSP operator produced a smaller standard deviation value of 9.04% (the same as LTP) which indicates that our proposed method does not only outperform the other operators on both datasets but also produced more consistent results.

## 7.2. Effects of Radius ( $R$ ) and Number of Neighbours ( $P$ )

In this section we are interested to know to what extent  $R$  and  $P$  affect the overall classification accuracy. For this purpose, we investigated seven different combination of  $R$  and  $P$  and tested each of them with all operators. Figures 21 and 22 show average maximum accuracies for both datasets using



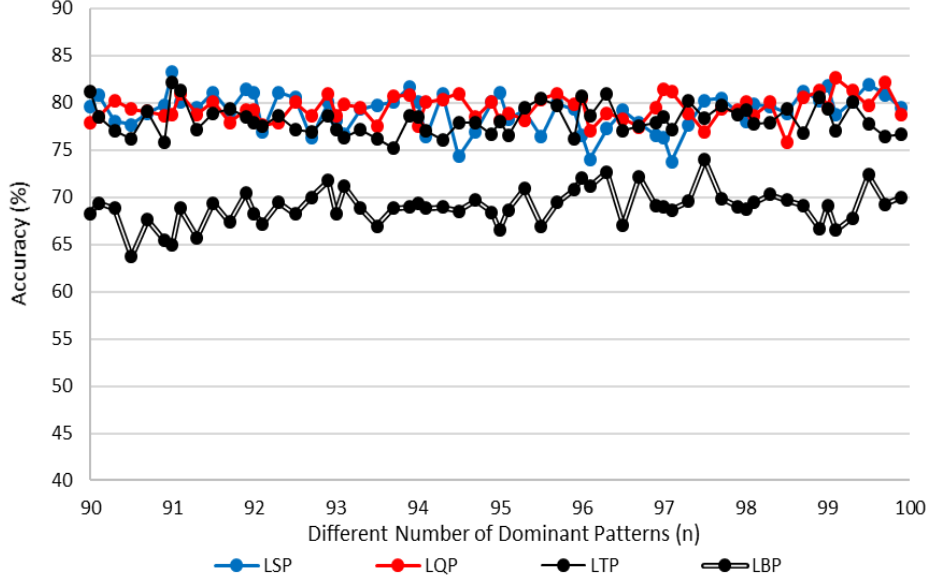


Figure 20: Effects of different numbers of dominant patterns ( $n$ ) across different operators ((P,R)=(18,9)) using SVM classifier on MIAS dataset.

the SVM classifier. For the LBP operator, local patterns extracted using smaller radii and less number of neighbours (e.g., (5, 10) and (5, 12)) tend to produce better results whereas the LTP operator produced its highest average maximum accuracy at (7, 14) ( $Acc > 78\%$ ). Similar results of 77.5% are obtained when the following parameters are employed: (7, 16), (9, 18) and (9, 20). However, when evaluated on the InBreast dataset, the highest average maximum accuracy (76%) is achieved at (9, 18) followed by parameters (7, 14) and (7, 16). For the LQP operator, the highest average maximum accuracy ( $Acc > 78\%$ ) is achieved at parameters (7, 14) followed by (9, 18) with only 0.5% difference. In the InBreast dataset, the LQP operator produced highest average maximum accuracy of 77.5% with parameters (9, 18). The LSP operator produced consistent average maximum accuracy when parameters (7, 14), (7, 16), (9, 18), (9, 20) and (11, 22) are used which is around 77%. Nevertheless, it produces at least 2.5% higher average maximum accuracy at (9, 20) compared to (7, 14), (7, 16), (9, 18) and (11, 22). These results indicate that when choosing the values for  $P$  and  $R$ , the following values (7, 16), (9, 18), (9, 20) are a good starting point. The LSP operator produced its best

599 results at (9, 20) using the SVM classification approach.

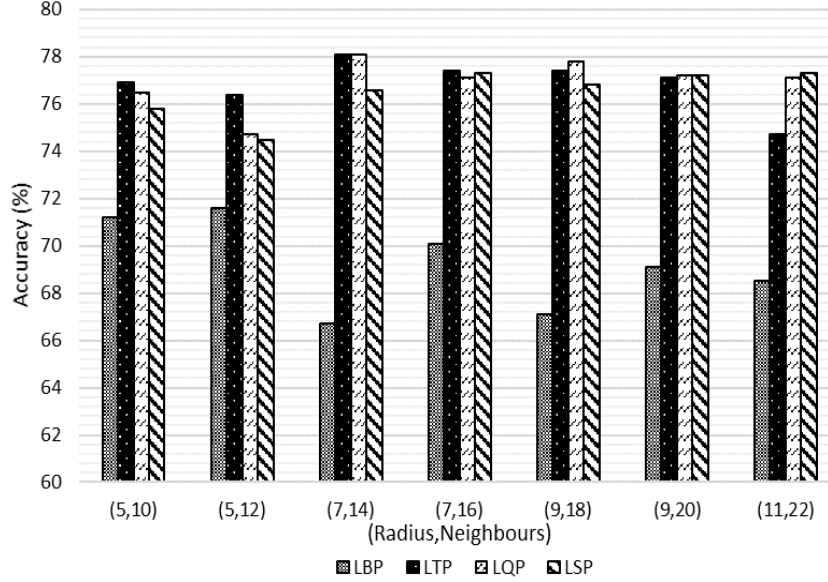


Figure 21: Average maximum accuracy (MIAS dataset) across different classifiers based on different operators and their associated parameters.

### 600 7.3. Why do multichannel local patterns work?

601 To answer this question we conducted several experiments by performing  
602 channel based classifications on the MIAS dataset. In the case of the  
603 LTP operator (three channels with two binary patterns), firstly we perform  
604 individual classification based on features extracted from the first binary  
605 patterns and secondly we concatenate features extracted from two or more  
606 channels. Table 3 shows our experimental results for the LTP, LQP and LSP  
607 operators. Note that the following parameters are used based on their best  
608 performance using the SVM classifier (see Figures 12, 13 and 14):  $LTP(7, 16)$   
609 with  $n = 93.5$ ,  $LBP(9, 18)$  with  $n = 99.9$  and  $LSP(9, 18)$  with  $n = 99.1$ .

610 It can be observed that classification based on features extracted from  
611 a single channel alone is insufficient. For the LTP operator, each chan-  
612 nel produced just over 73% but concatenating local patterns from all chan-  
613 nels produced 81%, yielding over 7% improvement. A similar case occurred  
614 for the LQP operator where binary patterns from a single channel pro-  
615 duced  $Acc < 70\%$ . However, combining binary patterns from the first three

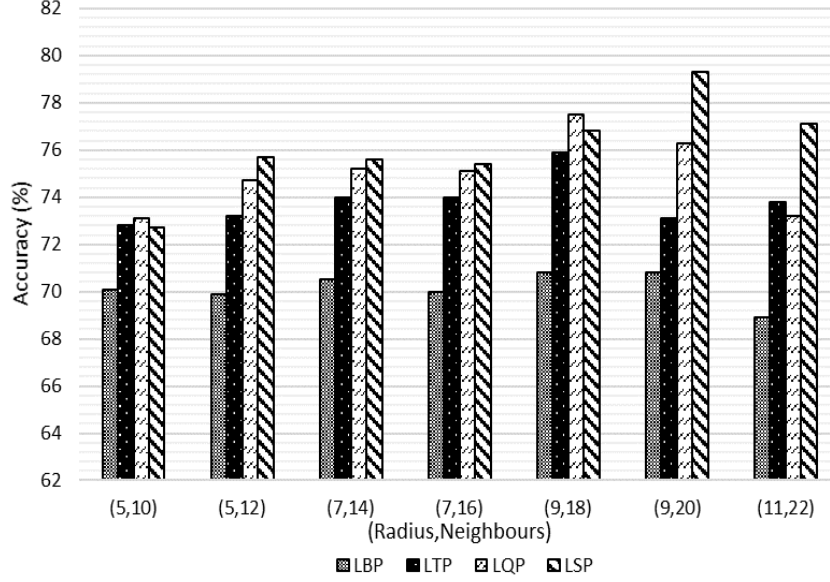


Figure 22: Average maximum accuracy (InBreast dataset) across different classifiers based on different operators and their associated parameters.

channels ( $Ch1 \oplus Ch2 \oplus Ch3$ ) yields a significant improvement 76.4% and concatenating all local patterns from all channels produced a maximum accuracy of 82.1%. For our proposed operator,  $Ch2$  and  $Ch5$  produced classification accuracy under 60% and  $Ch1$ ,  $Ch3$  and  $Ch4$  produced accuracy close to 70%. Concatenating all features from all channels resulted in  $Acc = 83.3\%$  however removing a less informative channel (e.g.,  $Ch2$ ) by considering  $Ch1 \oplus Ch3 \oplus Ch4 \oplus Ch5 \oplus Ch6$ , the classification accuracy improved to 83.8%.

Based on these experiments we found that encoding an image by dividing it into several channels captures more textural details which is difficult to capture with the original LBP operator. This can be clearly seen in Figures 8 and 9 where most accuracies achieved by the LBP operator are under 70% whereas the other operators consistently achieved above 75% accuracy. This also indicates that the LQP and LTP operators capture more details of the Corpus Mammar region in comparison to the LBP operator. Similarly, the LSP operator captures even more details compared to LQP and LTP operators. However, a higher order encoding system such as dividing into nine or eleven channels does not necessarily improve the classification accuracy

Operators	Channel ( $Ch$ )	Channels Concatenation ( $\oplus$ )
LTP	$Ch1 = 73.3 \pm 8.6$ $Ch2 = 74.1 \pm 9.1$	$Ch1 \oplus Ch2 = 81.95$
LQP	$Ch1 = 59.8 \pm 10.6$ $Ch2 = 65.9 \pm 9.6$ $Ch3 = 68.9 \pm 11.5$ $Ch4 = 66.5 \pm 9.9$	$Ch1 \oplus Ch2 = 69.3 \pm 8.3$ $Ch1 \oplus Ch2 \oplus Ch3 = 76.4 \pm 8.1$ $Ch1 \oplus Ch2 \oplus Ch3 \oplus Ch4 = 82.1 \pm 7.1$
LSP	$Ch1 = 68.9 \pm 9.4$ $Ch2 = 52.6 \pm 11.6$ $Ch3 = 70.1 \pm 8.8$ $Ch4 = 69.1 \pm 8.9$ $Ch5 = 59.3 \pm 11.3$ $Ch6 = 65.7 \pm 12.8$	$Ch1 \oplus Ch2 = 74.3 \pm 7.9$ $Ch1 \oplus Ch2 \oplus Ch3 = 75.1 \pm 7.9$ $Ch1 \oplus Ch2 \oplus Ch3 \oplus Ch4 = 70.1 \pm 10.1$ $Ch1 \oplus Ch2 \oplus Ch3 \oplus Ch4 \oplus Ch5 = 77.7 \pm 9.1$ $Ch1 \oplus Ch3 \oplus Ch4 \oplus Ch6 = 80.7 \pm 10.1$ $Ch1 \oplus Ch3 \oplus Ch4 \oplus Ch5 \oplus Ch6 = 83.8 \pm 9.8$ $Ch1 \oplus Ch2 \oplus Ch3 \oplus Ch4 \oplus Ch5 \oplus Ch6 = 83.3 \pm 8.8$

Table 3: Classification results (%) based on binary patterns extracted from individual channel (second column) and concatenated binary patterns from two or more channels

because it increases the number of features yielding more complex decision boundaries. An early indication can be observed based on our experiment results in Table 3 where best accuracy can be achieved by the LSP operator when excluding local patterns from  $Ch2$ . This reduces the number of features yielding a less complex decision boundaries in a feature space, hence boosting the overall performance of the method.

A channel encoding system also extracts a set of ‘weak’ features which produce satisfactory results on its own, but when combined they provide good results (similar to ensemble classifiers such as the Random Forests). For example, in Table 3 combining two or more local patterns from different channels always produce better results. This is due to each channel containing unique information which is not available in the other channels and they complement each other when combined. For example, individual classification in LQP produce under 70% accuracy but improve to over 76% and 82% when three and four channels are combined, respectively. Simi-

larly, all channels from the LSP operator are unable to provide sufficient information of the Corpus Mammæ region individually but when combining  $Ch1 \oplus Ch3 \oplus Ch4 \oplus Ch5 \oplus Ch6$  we achieved 83.8% accuracy.

Moreover, the threshold values which are determined automatically enable the operator to capture robust texture information as pixel information is encoded according to the local threshold values rather than a global threshold value as in the LTP and LQP operators. Local threshold values enable the LSP operator to capture textural information based on the topology and intensity distribution of the image and hence are more discriminant and tolerant to noise. In contrast, a global threshold value does not consider intensity distribution of a region. As a result, it may affect the discriminatory level of the local patterns.

#### 7.4. Extending to 9-encoding and 11-encoding systems

To further evaluate the performance of different encoding systems, we conducted several experiments by extending the LSP to 9- and 11-encoding systems both on the MIAS and InBreast datasets. Figures 23, 24, 25 and 26 show that the performance is decreasing as we use 9- and 11-encoding systems.

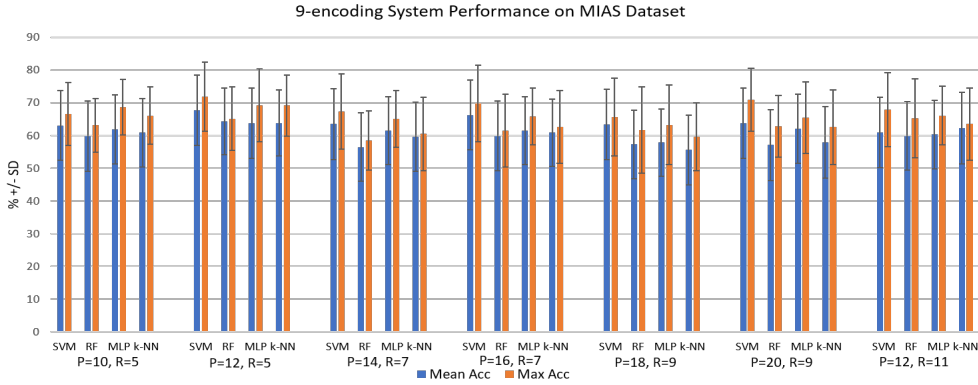


Figure 23: 9-encoding system performance on MIAS dataset

Experimental results suggest that the 7-encoding system the LSP achieved the best performance and eventually decreased to between 55% to 63%. From our own observation, the main reason for this is due to higher encoding systems (e.g. 9-encoding) producing very sparse features (containing many

671 zeros). This is similar when performing the 11-encoding system (in fact fea-  
 672 tures are even sparser). The main problem with sparse features is they are  
 673 most likely being ignored by most classifiers particularly tree-based machine  
 674 learning algorithms. Secondly, in most cases zeros are less informative mak-  
 675 ing the features less meaningful and hence less discriminant. As a result,  
 676 performance classification reduced significantly. In Figure 25 and 26 none of  
 677 the maximum accuracies achieved was above 70% whereas the LBP operators  
 678 achieved above this value in some cases.

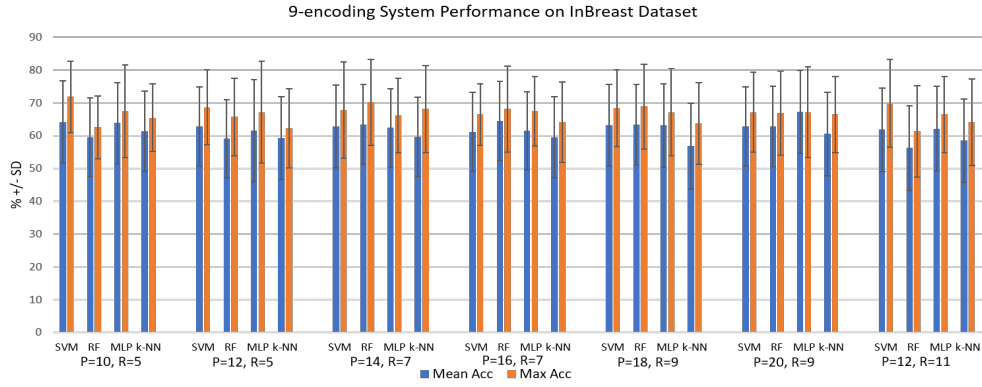


Figure 24: 9-encoding system performance on InBreast dataset

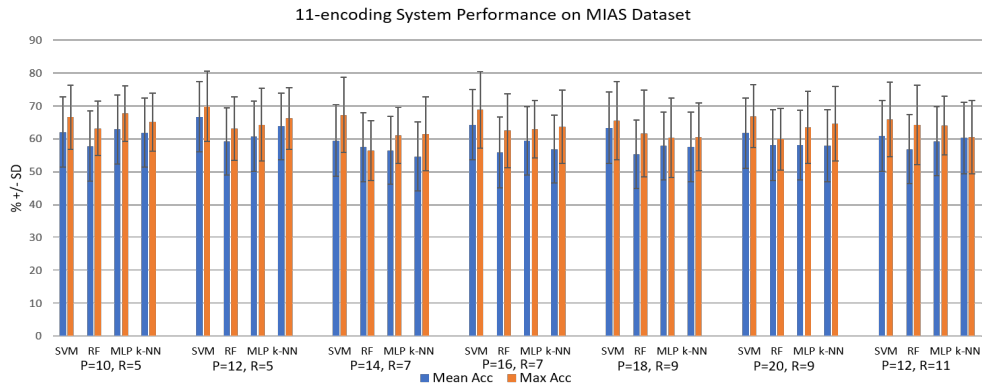


Figure 25: 11-encoding system performance on MIAS dataset

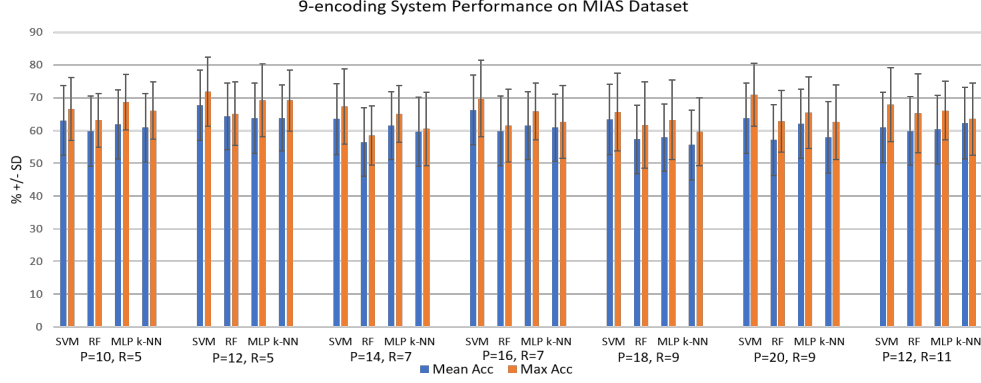


Figure 26: 11-encoding system performance on InBreast dataset

#### 679 7.5. Existing Methods in the Literature

680 The main goal of our study is to improve the performance of binary-  
681 based local pattern feature extraction methods, namely Local Binary Pat-  
682 tern (LBP), Local Ternary Pattern (LTP) and Local Quinary Pattern (LQP).  
683 Our study does not intend to quantitatively compare the performance of the  
684 LSP operator against other feature extraction methods. There are many fea-  
685 ture extraction methods in the literature, therefore quantitative comparison  
686 is extremely difficult. It should also be noted that quantitative compari-  
687 son with the other breast density methods in the literature is because most  
688 studies combine features from different feature extraction algorithms such as  
689 first-order statistical features, second-order statistical features, morphologi-  
690 cal features, features of Gabor filters, edge information, etc.

691 Many breast density classification methods have been developed over the  
692 last two decades. The best classification accuracy (based on BI-RADS classes  
693 (fourth edition)) reported in the literature is over 90% by the study by [14]  
694 followed by the study of [11] with 86% classification accuracy. However, these  
695 methods are computationally expensive due to different numbers of feature  
696 extraction techniques employed. Several techniques used to extract different  
697 types of features such as the first order statistical techniques to compute the  
698 local pixel intensity, the co-occurrence matrix to extract grey level distribu-  
699 tion, an adaptive thresholding technique to extract the region's morphological  
700 information, Gabor filter to obtain edges information, etc. Also, their meth-  
701 ods employed the Fuzzy C-Means clustering technique as a separate process  
702 to classify the breast tissue into different pixel intensities. Once again this

process is time consuming especially when dealing with large images such as mammograms. In contrast, the LSP operator (with accuracy close to 84%) extracts and processes all this information at once, hence reducing the computational complexity of the system.

Previously, [20] used the LTP operator to extract local information and achieved a promising result of 82.33% accuracy. However, this method suffers from having to deal with a large number of features (over one thousand) due to the multi-orientation approach (e.g., ten histograms from ten orientations concatenated). Our recent study, [21] reported the best accuracy of over 86% using multi-orientation LQP operator and combining other local patterns extracted using on ellipse neighbourhood. Nevertheless, this method contains several separate stages (e.g. extraction using different neighbourhood topologies (circle and ellipse), different orientations of resolutions which can be tedious whereas the LSP operator uses only one neighbourhood topology (e.g. a circle) with a single resolution. Other studies in the literature such as [11, 12, 13, 14, 15, 18, 19, 22, 23] reported accuracies under 80%. However, note that the purpose of our study is not to develop a breast density classification method that can necessarily outperform the other methods in the literature but to study the effects of various channel encoding techniques in the original LBP operator which have lead us to a more robust technique called the LSP operator, improving the performance of the LBP, LTP and LQP operators.

## 7.6. Statistical Analysis

We performed statistical analysis to investigate whether there is a significant difference/improvement at  $p \leq 0.05$  between the best results achieved by the LSP operator and the other operators discussed in this study on both datasets. For this purpose, we compared the best maximum (Max) and average (Avg) accuracies produced by the LSP with the best results of LBP, LTP and LQP operators. The  $p$ -value was computed using a  $t$ -test comparing each result of each operator with the best result achieved by the LSP operator according to their respective metric (e.g., Max *Acc* or Avg *Acc*). The size of population for each dataset are 103 (206 images) and 161 (322 images) for the InBreast and MIAS dataset, respectively.

Table 4 shows the  $p$ -values between the best accuracies achieved by the LSP operator and the other operators for both datasets. In terms of maximum accuracy, there is a significant improvement at  $p \leq 0.05$  when using the LSP features in classifying breast density on the MIAS ( $p = 0.0015$ ) and



Dataset	Results	LBP	LTP	LQP	LSP
MIAS	Max <i>Acc</i> (%)	73.8 $\pm$ 10.6	81 $\pm$ 9.5	82.1 $\pm$ 7.1	83.3 $\pm$ 8.8
	( <i>p</i> value)	(0.0015)	(0.2260)	(0.4019)	-
	Avg <i>Acc</i> (%)	69.7 $\pm$ 9.7	77.5 $\pm$ 9.2	78.6 $\pm$ 8.3	81.6 $\pm$ 9.1
	( <i>p</i> value)	(0.00001)	(0.0428)	(0.0433)	-
InBreast	Max <i>Acc</i> (%)	73.8 $\pm$ 9.9	78.7 $\pm$ 11.2	80.1 $\pm$ 10.5	80.5 $\pm$ 9.2
	( <i>p</i> value)	(0.0064)	(0.2646)	(0.4489)	-
	Avg <i>Acc</i> (%)	70.2 $\pm$ 11.7	74.3 $\pm$ 11.4	75.5 $\pm$ 11.3	77.1 $\pm$ 10.9
	( <i>p</i> value)	(0.0007)	(0.0137)	(0.2151)	-

Table 4: Classification results (%) and *p*-values between the best accuracies achieved by the LSP operator and the other operators discussed in this study.

740 InBreast ( $p = 0.0007$ ) datasets. Nevertheless, the best maximum results of  
741 the LSP operator are not statistically significant at  $p \leq 0.05$  in compari-  
742 son with the best accuracies achieved by the LTP and LQP operators. The  
743 *p*-values are 0.2260 and 0.4019 for the LTP and LQP operators, respectively  
744 when evaluated on the MIAS dataset which are similar when tested on the  
745 InBreast dataset,  $p = 0.2646$  and  $p = 0.4489$  for the LTP and LQP oper-  
746 ators, respectively. The results are statistically significant at  $p \leq 0.05$  on  
747 the MIAS dataset when comparing the best average accuracy produced by  
748 the LSP features ( $Acc = 81.6\%$ ) with the other features with *p*-values are  
749 0.00001, 0.0428 and 0.0433 for LBP, LTP and LQP respectively. Similarly,  
750 when tested on the InBreast dataset the best average accuracy produced by  
751 the LSP operator is significantly better in comparison to the results of LBP  
752 ( $p = 0.0007$ ) and LTP ( $p = 0.0137$ ) operators. However, the best average  
753 accuracy of LSP is not significant ( $p = 0.2151$ ) compared to the result of the  
754 LQP operator.

### 755 7.7. Limitations of our study

756 We highlight the following limitations of our study:

- 757 1. The sample dataset used to optimise the classifiers' parameters was  
758 taken from the MIAS dataset which only 8-bit contrast. This means  
759 the optimised parameters and testing results might be different when

testing with real clinical data (12 to 14 bits contrast). The reasons we chose the MIAS dataset rather than the InBreast dataset when optimising the classifiers' parameters are (a) it has larger number of images/cases hence giving us a larger number of test cases and (b) the number of cases for each class is more representative compared to the number of cases for each class in the InBreast dataset. For example, the number images of BI-RADS IV is 14 which is extremely small. Including cases in our dataset for parameter optimization, will make our testing dataset extremely imbalanced. Although in this study we used images from the MIAS dataset for parameter optimization, in a real clinical environment this procedure can be easily changed/adapted by using a sub-sample from the new dataset.

2. The MIAS dataset is a somewhat old database and hence does not represent the actual contrast representation of the latest mammograms. We included the MIAS dataset in our study due to (a) it is the most used and compared dataset in the literature because of its availability whereas other datasets are not easily accessible, (b) acquiring mammogram datasets is very difficult especially the ones with BI-RADS (version 4) classification and (c) other datasets do not classify each breast image based on BI-RADS density but are classified based on a small square patch based on tissue types such as 'fatty', 'glandular' or 'breast tissue'.
3. The parameters  $P$ ,  $R$  and  $n$  were tested/selected empirically. Nevertheless, our testing strategy is based on the previous studies of Ojala et al. [32], Tan and Triggs [33], Nanni et al. [34]. To the best of our knowledge this a common approach used in the literature especially for parameters  $P$  and  $R$ . For parameter  $n$ , we tested 60 different values from  $n = 90$  to  $n = 99.9$  with 0.2 interval (e.g. 90, 90.1, 90.3, 90.5, etc). This should be sufficient to investigate the performance variations across different  $n$  values.
4. The number of cases used in the development of parameter optimization is small. However, we prefer to optimise the number of images for testing purposes and therefore, we limit the number of cases for parameter optimisation. It is more important to test the model performance on a larger size of dataset rather than on a smaller dataset.

### 7.8. Future Work

For future work we plan to consider the following:

1. Since multichannel based operators such as LTP, LQP and LSP require the number of dominant patterns ( $n$ ) to obtain an optimum classification accuracy, we plan to develop a statistical method that can automatically determine an optimum value of  $n$ .
2. We also interested to investigate the performance of LBP's two-channel variants such as M-ELBP, ELBP, MRELBP, STLBP and CLBP when multichannel based operators are being applied.
3. As indicated in Table 3, combining all local patterns from all channels does not necessarily produce optimal accuracy. For example, the LSP operator achieved its highest accuracy when excluding *Ch2*. Therefore, we plan to develop a method that can automatically determine the most informative local patterns from each channel and in the classification phase only combine the most discriminant features. This can be done by measuring mutual information among the histogram features and combining the ones with the least overlapping information.
4. We plan to develop a deep learning based network to extract local features of the Corpus Mammarum region and combine them with features extracted with the LQP or LSP operators.

## 8. Summary and Conclusions

In this paper, we studied the effects of various channel encoding techniques in local pattern extraction which have led us to a more robust technique called the LSP operator. We investigated the following aspects in our study:

1. We made comparisons between the LBP, LTP, LQP and LSP operators based on MIAS and InBreast datasets. Experimental results suggest that the proposed texture operator outperformed the other operators on both datasets.
2. We investigated whether the choice of a classifier can influence the performance of the system. We employed four different classifiers and found that in many cases the SVM classifier outperforms the other classifiers. Local patterns extracted with the LTP operator tend to produce better results when the MLP or  $k$ -NN classifier is employed whereas the SVM classifier produced better accuracy using LQP and LSP features.

- 831 3. We conducted experiments to investigate whether the operators dis-  
832 cussed in this study have a particular preference regarding  $R$  and  $P$ .  
833 We found that in many cases LBP features produced similar results of  
834 under 70% accuracy. However, the LSP features tend to produce better  
835 classification results when  $LTP(7, 14)$  and  $LTP(9, 18)$  are used. On the  
836 other hand, both LQP and LSP operators tend to produce consistent  
837 results when  $R \in \{7, 9\}$  and  $P \in \{16, 18, 20\}$  are used.
- 838 4. We studied the effects of selecting different numbers of dominant pat-  
839 terns and experimental results show that the value of  $n$  plays an impor-  
840 tant role in obtaining optimal accuracy. The LSP and LTP operators  
841 produced variation accuracy of 9.04% compared to 9.25% and 9.71%  
842 for LQP and LBP, respectively.
- 843 5. We also studied channel based classification and concatenating local  
844 patterns from different channels. Experimental results suggest that  
845 each channel contains unique features which are not available in the  
846 other channels and combining these features yielded to better classifi-  
847 cation results. We also found that removing the most uninformative  
848 channel in LSP features improves the classification accuracy by 0.5%.
- 849 6. We developed simple methods to automatically determine the threshold  
850 values in LSP by computing the 25<sup>th</sup> percentile and 75<sup>th</sup> percentile of  
851 the neighbouring pixels.

852 In conclusion, we have studied various channel encoding techniques in  
853 LBP, LTP and LQP operators where we found that multichannel local pat-  
854 terns are more robust in discriminating different classes of breast density.  
855 This study has led us to the development of a seven-encoding technique call  
856 LSP operator to capture more texture details within the Corpus Mammae re-  
857 gion. We also introduced a simple method to automatically determine the  
858 threshold values in the LSP operator by computing the first order statisti-  
859 cal values of the neighbourhood pixels. Experimental results show that the  
860 LSP features outperformed the LBP, LTP and LQP operators on both MIAS  
861 and InBreast datasets. The LSP features produced a maximum accuracy of  
862 83.8% using the SVM classifier when combining local patterns from  $Ch1$ ,  
863  $Ch3$ ,  $Ch4$ ,  $Ch5$  and  $Ch6$  on the MIAS dataset. The LBP, LTP and LQP  
864 features achieved maximum accuracies of 73.8%, 81% and 82.1%, respec-  
865 tively. Based on these experimental results, the proposed seven-encoding  
866 system approach (LSP) is shown to be a robust and more consistent texture  
867 operator.

## 868 Declaration of competing interest

869 None Declared.

## 870 Acknowledgment

871 This research was undertaken as part of the Decision Support and In-  
872 formation Management System for Breast Cancer (DESIREE) project. The  
873 project has received funding from the European Union's Horizon 2020 re-  
874 search and innovation programme under grant agreement No 690238.

## 875 References

- 876 [1] American Cancer Society, Breast cancer facts and figures  
877 2017/2018, 2018. Available at [https://www.cancer.org/content/  
878 dam/cancer-org/research/cancer-facts-and-statistics/  
879 breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.  
880 pdf](https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf) Accessed: 2017-06-23.
- 881 [2] C. E. DeSantis, J. Ma, A. G. Sauer, L. A. Newman, A. Jemal, Breast  
882 cancer statistics, 2017, racial disparity in mortality by state., CA: A  
883 Cancer Journal for Clinicians 67 (2017) 439–448.
- 884 [3] Breast Cancer, UK, Breast cancer care. facts and statistics 2018,  
885 2018. Available at [https://www.breastcancercare.org.uk/  
886 about-us/media/press-pack-breast-cancer-awareness-month/  
887 facts-statistics](https://www.breastcancercare.org.uk/about-us/media/press-pack-breast-cancer-awareness-month/facts-statistics) Accessed: 2017-05-23.
- 888 [4] Breast Cancer, UK, Breast cancer symptoms, 2018. Avail-  
889 able at [https://www.cancerresearchuk.org/about-cancer/  
890 breast-cancer/symptoms](https://www.cancerresearchuk.org/about-cancer/breast-cancer/symptoms) Accessed: 2017-06-26.
- 891 [5] Breast Cancer, U.S. breast cancer statistics, 2016. Available at [http:  
892 //www.breastcancer.org/symptoms/understand\\_bc/statistics](http://www.breastcancer.org/symptoms/understand_bc/statistics)  
893 Accessed: 2017-06-28.
- 894 [6] Breastcancer.org, Dense breast, 2018. Available at [https://www.  
895 breastcancer.org/risk/factors/dense\\_breasts](https://www.breastcancer.org/risk/factors/dense_breasts) Accessed: 2017-06-  
896 28.

- 897 [7] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi,  
898 M. Ghafoorian, J. A. van der Laak, B. van Ginneken, C. I. Sánchez,  
899 A survey on deep learning in medical image analysis, *Med. Image Anal.*  
900 42 (2017) 60–80.
- 901 [8] A. Hamidinekoo, E. Denton, A. Rampun, K. Honnor, R. Zwiggelaar,  
902 Deep learning in mammography and breast histology, an overview and  
903 future trends, *Med Image Anal* 47 (2018) 45–67.
- 904 [9] N. F. Boyd, J. W. Byng, R. A. Jong, E. K. Fishell, L. E. Little, A. B.  
905 Miller, G. A. Lockwood, D. L. Tritchler, M. J. Yaffe, Quantitative  
906 classification of mammographic densities and breast cancer risk: Results  
907 from the canadian national breast screening study, *J. Nat. Cancer Inst.*  
908 87 (1995) 670–675.
- 909 [10] J. Suckling et al., The mammographic image analysis society digital  
910 mammogram database, in: *Proc. Excerpta Med. Int. Congr. Ser.*, pp.  
911 375–378.
- 912 [11] A. Oliver, J. Freixenet, R. Martí, E. P. J. Pont, E. R. E. Denton,  
913 R. Zwiggelaar, A novel breast tissue density classification methodol-  
914 ogy, *IEEE Trans. Inf. Technol. Biomed.* 12 (2008) 55–65.
- 915 [12] K. Bovis, S. Singh, Classification of mammographic breast density using  
916 a combined classifier paradigm, in: *Proc. Int. Workshop on Digital*  
917 *Mammography, Lecture Notes in Computer Science*, Springer, 2002, pp.  
918 177–180.
- 919 [13] M. Muštra, M. Grgić, K. Delač, Breast density classification using mul-  
920 tiple feature selection, *Automatika* 53 (2012) 362–372.
- 921 [14] N. M. Parthaláin, R. Jensen, Q. Shen, R. Zwiggelaar, Fuzzy-rough  
922 approaches for mammographic risk analysis, *Intell. Data Anal.* 14 (2010)  
923 225–244.
- 924 [15] D. Raba, J. Martí, R. Martí, M. Peracaula, Breast mammography asym-  
925 metry estimation based on fractal and texture analysis, in: *Proc. Com-*  
926 *puted Aided Radiology and Surgery*, Berlin, Germany, volume 1398.

- 927 [16] H. Li, M. L. Giger, O. I. Olopade, A. Margolis, L. Lan, M. R. Chinander,  
928 Computerized texture analysis of mammographic parenchymal patterns  
929 of digitized mammograms, *Acad. Radiol.* 12 (2005) 863–73.
- 930 [17] A. Bosch, X. Munoz, A. Oliver, J. Martí, Modeling and classifying  
931 breast tissue density in mammograms, in: *Proc IEEE Comput. Soc.  
932 Conf. Comput. Vis. Pattern. Recognit. (CVPR)*, volume 2, pp. 1552–  
933 1558.
- 934 [18] Z. Chen, E. Denton, R. Zwigelaar, Local feature based mammographic  
935 tissue pattern modelling and breast density classification, in: *Proc.  
936 4th International Conference on Biomedical Engineering and Informatics  
937 (BMEI)*, volume 1, pp. 351–355.
- 938 [19] S. Petroudi, T. Kadir, M. Brady, Automatic classification of mammo-  
939 graphic parenchymal patterns: A statistical approach, in: *Proc. IEEE  
940 Conf. Eng. Med. Biol. Soc.*, volume 1, pp. 798–801.
- 941 [20] A. Rampun, P. J. Morrow, B. W. Scotney, R. J. Winder, Breast density  
942 classification using local ternary patterns in mammograms, in: *Proc. Im-  
943 age Analysis and Recognition*, Springer International Publishing, Cham,  
944 2017, pp. 463–470.
- 945 [21] A. Rampun, B. W. Scotney, P. J. Morrow, H. Wang, J. Winder, Breast  
946 density classification using local quinary patterns with various neigh-  
947 bourhood topologies, *J. Imaging* 4 (2018) 14.
- 948 [22] M. George, A. Rampun, E. Denton, R. Zwigelaar, Mammographic  
949 ellipse modelling towards birads density classification, in: *Proc. Inter-  
950 national Workshop on Breast Imaging (IWDM) 2016: Breast Imaging*,  
951 pp. 423–430.
- 952 [23] M. George, R. Zwigelaar, Breast tissue classification using local binary  
953 pattern variants: A comparative study (in press), in: *Proc. International  
954 Workshop on Breast Imaging (IWDM) 2016: Breast Imaging*.
- 955 [24] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Méridaa, C. I.  
956 Sánchez, R. Manna, A. den Heeten, N. Karssemeijer, Large scale deep  
957 learning for computer aided detection of mammographic lesions, *Med.  
958 Image Anal.* 35 (2017) 303–312.

- 959 [25] A. Rampun, K. López-Linares, P. J. Morrow, B. W. Scotney, H. Wang,  
960 I. G. Ocaña, G. Maclair, R. Zwigelaar, M. A. González Ballester,  
961 I. Macía, Breast pectoral muscle segmentation in mammograms using  
962 a modified holistically-nested edge detection network, *Medical Image*  
963 *Analysis* 57 (2019) 1–17.
- 964 [26] A. A. Mohamed, W. A. Berg, H. Peng, Y. Luo, R. C. Jankowitz, S. Wu.,  
965 A deep learning method for classifying mammographic breast density  
966 categories, *Med. Phys.* 45 (2017) 314–321.
- 967 [27] S. Li, J. Wei, C. Heang-Ping, M. A. Helvie, M. A. Roubidoux, Y. Lu,  
968 C. Zhou, L. M. Hadjiiski, R. K. Samala, Computer-aided assessment  
969 of breast density: comparison of supervised deep learning and feature-  
970 based statistical learning, *Phys. Med. Biol.* 63 (2018) 15.
- 971 [28] M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel,  
972 C. M. Vachon, K. Holland, R. R. Winkel, N. Karssemeijer, M. Lill-  
973 holm, Unsupervised deep learning applied to breast density segmenta-  
974 tion and mammographic risk scoring, *IEEE Trans. Med. Imaging.* 35  
975 (2016) 1322–1331.
- 976 [29] J. Lee, R. M. Nishikawa, Automated mammographic breast density  
977 estimation using a fully convolutional network, *Med. Phys.* 45 (2018)  
978 1178–1190.
- 979 [30] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, J. S.  
980 Cardoso, Inbreast: toward a full-field digital mammographic database,  
981 *Acad. Radiol.* 19 (2011) 236–428.
- 982 [31] A. Rampun, P. J. Morrow, B. W. Scotney, R. J. Winder, Fully auto-  
983 mated breast boundary and pectoral muscle segmentation in mammo-  
984 grams, *Artif. Intell. Med.* 79 (2017) 28–41.
- 985 [32] T. Ojala, M. Pietikainen, T. Maenpaar, Multiresolution gray-scale and  
986 rotation invariant texture classification with local binary patterns, *IEEE*  
987 *Trans. Pattern Anal. Mach. Intell.* 24 (2002) 971–987.
- 988 [33] X. Tan, B. Triggs, Enhanced local texture feature sets for face recogni-  
989 tion under difficult lighting conditions, *IEEE Trans. Image Process.* 19  
990 (2010) 1635–1650.



- 991 [34] L. Nanni, A. Luminia, S. Brahnem, Local binary patterns variants as  
992 texture descriptors for medical image analysis, *Artif. Intell. Med.* 49  
993 (2010) 117–125.
- 994 [35] L. Liu, L. Zhao, Y. Long, G. Kuang, P. Fieguth, Extended local binary  
995 patterns for texture classification, *Image Vision Comput.* 30 (2012) 86  
996 – 99.
- 997 [36] S. Liao, A. C. S. Chung, Face recognition by using elongated local  
998 binary patterns with average maximum distance gradient magnitude,  
999 in: *Proc. Computer Vision—ACCV 2007*, volume 4844, Lecture Notes  
1000 in Computer Science, 2007, pp. 672–679.
- 1001 [37] L. Liu, S. Lao, P. W. Fieguth, Y. Guo, X. Wang, M. Pietikäinen, Median  
1002 robust extended local binary pattern for texture classification, *IEEE*  
1003 *Trans. Image Process.* 25 (2016) 1368–1381.
- 1004 [38] J. Yang, S. Wang, Z. Lei, Y. Zhao, S. Z. Li, Spatio-temporal lbp based  
1005 moving object segmentation in compressed domain, in: *Proc. IEEE*  
1006 *Ninth International Conference on Advanced Video and Signal-Based*  
1007 *Surveillance*, pp. 252–257.
- 1008 [39] Z. Guo, L. Zhang, D. Zhang, A completed modeling of local binary  
1009 pattern operator for texture classification, *IEEE Trans. Image Process.*  
1010 19 (2010) 1657–1663.
- 1011 [40] Y. Gio, G. Zhao, M. Pietikäinen, Discriminative features for feature  
1012 description, *Pattern Recognit.* 45 (2012) 3834–3843.
- 1013 [41] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Wit-  
1014 ten, The weka data mining software: an update, *ACM SIGKDD explo-*  
1015 *rations newsletter* 11 (2009) 10–18.